Heriot-Watt University

PhD Thesis

# 3D reconstruction and object recognition from 2D SONAR data

*Author:*
Thomas Guerneve

*Supervisors:*
Prof. Yvan Petillot
Dr. Kartic Subr

*A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy*

*in the*

School of Engineering and Physical Sciences

August 2018

# *Abstract*

Accurate and meaningful representations of the environment are required for autonomy in underwater applications. Thanks to favourable propagation properties in water, acoustic sensors are commonly preferred to video cameras and lasers but do not provide direct 3D information. This thesis addresses the 3D reconstruction of underwater scenes from 2D imaging SONAR data as well as the recognition of objects of interest in the reconstructed scene. We present two 3D reconstruction methods and two model-based object recognition methods. We evaluate our algorithms on multiple scenarios including data gathered by an AUV. We show the ability to reconstruct underwater environments at centimetre-level accuracy using 2D SONARs of any aperture. We demonstrate the recognition of structures of interest on a medium-sized oil-field type environment providing accurate yet low memory footprint semantic world models. We conclude that accurate 3D semantic representations of partially-structured marine environments can be obtained from commonly embedded 2D SONARs, enabling online world modelling, relocalisation and model-based applications.

# *Acknowledgements*

I wish to express my gratitude to Professor Yvan Petillot, my first supervisor, who introduced me to the fascinating world of research and gave me the opportunity to start this adventure. Undoubtedly, his guidance throughout the three years as well as his human qualities were decisive in making this experience a success.

I would like to thank Dr. Kartic Subr for his patient guidance. The numerous valuable discussions we had made this experience truly enriching. His expertise and support were precious and greatly appreciated in crucial times of this project.

My sincere thanks to the personal of SeeByte for creating a great work environment. A special thanks goes to the AIV team and in particular Chris Sotzing, Gavin Irvine, Mark Payne and Brieuc Roblin as well as Subsea7 for giving me access to fantastic datasets and developing the coolest underwater vehicle in the world.

I would like to thank the members of the Ocean System Lab in Heriot Watt university for the many valuable exchanges. I am particularly grateful to Len McLean for his help during tank experiments and in-lake trials. Likewise, I would like to thank Yan Pailhas for his advices and support.

I had the pleasure to be part of Robocademy, a Marie Curie Research Training Network - FP7-PEOPLE-2013-ITN-608096. I am grateful to the management board of this project for making these collaborations possible. Not only this project provided valuable technical training but it also offered unrivalled opportunities to reach worldwide experts. Likewise, I would like to thank all the students and supervisors involved in this project for the all these good times spent (net)working together.

Finally, none of this would have been possible without the unconditional support of my family. I am deeply grateful for their constant encouragement and understanding.

ACADEMIC REGISTRY
**Research Thesis Submission**

| | | | |
|---|---|---|---|
| Name*:* | | | |
| School: | | | |
| Version: *(i.e. First, Resubmission, Final)* | | Degree Sought: | |

**Declaration**

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

| | | | |
|---|---|---|---|
| Signature of Candidate*:* | | Date: | |

**Submission**

| | |
|---|---|
| Submitted By *(name in capitals):* | |
| Signature of Individual Submitting: | |
| Date Submitted: | |

**For Completion in the Student Service Centre (SSC)**

| | | | |
|---|---|---|---|
| Received in the SSC by *(name in capitals):* | | | |
| *Method of Submission* *(Handed in to SSC; posted through internal/external mail):* | | | |
| *E-thesis Submitted (**mandatory for final theses)*** | | | |
| Signature: | | Date: | |

**Please note this form should be bound into the submitted thesis.**
Academic Registry/Version (1) August 2016

# Contents

# List of Tables

# List of Figures

# Abbreviations

| | |
|---|---|
| **3D** | 3 **D**imensions |
| **AIV** | **A**utonomous **I**nspection **V**ehicle |
| **AUV** | **A**utonomous **U**nderwater **V**ehicle |
| **CAD** | **C**omputer-**A**ided **D**esign |
| **CPU** | **C**entral **P**rocessing **U**nit |
| **DOA** | **D**irection **O**f **A**rrival |
| **DVL** | **D**oppler **V**elocity **L**og |
| **FLS** | **F**orward **L**ooking **S**onar |
| **FPGA** | **F**ield **P**rogrammable **G**ated **A**rray |
| **FOV** | **F**ield **O**f **V**iew |
| **FSPA** | **F**requency **S**teered **P**hased **A**rray |
| **GPS** | **G**lobal **P**ositinning **S**ystem |
| **GPU** | **G**raphics **P**rocessing **U**nit |
| **ICP** | **I**terative **C**losest **P**oint |
| **LIDAR** | **LI**ght **D**etection **A**nd **R**anging |
| **MAP** | **M**aximum **A** **P**osteriori |
| **NED** | **N**orth **E**ast **D**epth |
| **NNLS** | **N**on **N**egative **L**east **S**quare |
| **OBB** | **O**riented **B**ounding **B**ox |
| **OSL** | **O**cean **S**ystems **L**aboratory |
| **PSF** | **P**oint **S**pread **F**unction |
| **RAM** | **R**andom-**A**ccess **M**emory |
| **RANSAC** | **RAN**dom **SA**mpling **C**onsensus |
| **ROV** | **R**emotely **O**perated **V**ehicle |
| **SAS** | **S**ynthetic **A**perture **S**onar |

| | |
|---|---|
| **SLAM** | **S**imultaneous **L**ocalization **A**nd **M**apping |
| **SNR** | **S**ignal to **N**oise **R**atio |
| **SONAR** | **SO**und **N**avigation **A**nd **R**anging |
| **SVD** | (S)ingular **V**alue **D**ecomposition |
| **TV** | **T**otal **V**ariation |
| **TVG** | **T**ime **V**arying **G**ain |
| **USBL** | **U**ltra **S**hort **B**ase**L**ine |

# Symbols

| | |
|---|---|
| $t$ | time |
| $c$ | sound speed in water |
| $p_e(t)$ | emitted pulse |
| $Amp(t)$ | time-varying amplitude |
| $\Omega(t)$ | phase modulation term |
| $\delta(t)$ | Dirac function |
| $\bigtriangledown$ | gradient operator |
| $p$ | pressure field |
| $\mu$ | reflectivity coefficient |
| $r$ | range |
| $\theta$ | elevation angle |
| $\phi$ | azimuth angle |
| $A_v$ | vertical aperture |
| $A_h$ | horizontal aperture |
| $I_{Sonar}(t, \phi)$ | intensity measured at time $t$ and azimuth angle $\phi$ |
| $B(\theta)$ | angular attenuation of a transducer (beam pattern) |
| $V(t, \theta, \phi)$ | acoustic potential |
| $\theta_{3dB}$ | angle of mid-power attenuation |
| $R(r, \theta, \phi)$ | reflectivity at 3D point $(r, \phi, \theta)$ |
| $S(r, \theta, \phi)$ | surface indicator function |
| $\vec{v}_{\theta\phi}$ | direction of propagation of emitted wave front |
| $\vec{n}_{r\theta\phi}$ | normal at the surface of the scatterer at $(r, \phi, \theta)$ |
| $\wp(\phi_i, \theta_j)$ | ray on indexed azimuth and elevation angles $\phi_i, \theta_j$ |
| $U(a, b)$ | uniform distribution on the interval $[a, b]$ |
| $m(r, \phi)$ | mean intensity at range r and azimuth angle $\phi$ |

| | |
|---|---|
| $\sigma(r, \phi)$ | standard variation of the intensity at range r and azimuth angle $\phi$ |
| $\Theta_i$ | $i^{th}$ angular section along the vertical aperture |
| $\Delta_\theta$ | size of angular section along the vertical aperture |
| $\oslash$ | diameter |
| $\beta$ | vertical beam pattern |
| $\rho$ | reflectivity |
| $h$ | blurring kernel |
| $\mathcal{F}[]$ | Fourier transform |
| $L_p$ | $p$-norm |
| $\Delta_z$ | sampling period along the Z-axis |
| $N$ | number of input samples |
| $K$ | number of vertical subdivisions along the vertical aperture |
| $P$ | number of reconstructed points along the Z-axis |
| $I$ | vector of $N$ observations |
| $b$ | vector of $P$ estimated intensities |
| $A$ | imaging model matrix |
| $A^\star$ | approximate imaging model matrix |
| $A^\perp$ | normal-based imaging model matrix |
| $A^O$ | occlusion and normal-based imaging model matrix |
| $\lambda$ | regularization coefficient |
| $r_{model}$ | CAD model radius |
| $r_{subspace}$ | subspace radius |
| $T$ | sampling period |
| $\ell_{model}$ | CAD model length along its principal directions |
| $V_{model}$ | CAD model volume |
| $O_{model}$ | CAD model voxel occupancy on a regular grid |
| $\alpha_\gamma$ | score to test $\gamma$ |

# List of Publications

- Guerneve, T. and Petillot, Y. (2015). Underwater 3d reconstruction using blueview imaging sonar. In *OCEANS 2015-Genova*, pages 1–7. IEEE.
- Guerneve, T., Subr, K., and Petillot, Y. (2018). Three-dimensional reconstruction of underwater objects using wide-aperture imaging sonar. *Journal of Field Robotics*.
- Guerneve, T., Subr, K., and Petillot, Y. (2017b). Underwater 3d structures as semantic landmarks in sonar mapping. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pages 614–619. IEEE.
- Guerneve, T., Subr, K., and Petillot, Y. (2017a). Cad-model-based 3d video mapping. In *OCEANS–Anchorage, 2017*, pages 1–5. IEEE.

# Chapter 1

# Introduction

"Ocean is more ancient than the mountains, and freighted with the memories and the dreams of Time."

*The White Ship* - H.P. Lovecraft, 1919

## 1.1 Motivations

*"Freighted with the memories and dreams of Time"*, the Ocean encapsulates all the components of human life. From being the prime source of nourishment for life to playing a major role in the regulation of climate, it directly shaped the physicality of human beings. Since early times, its immensity gave birth to many mysteries, cultivating both human fears and dreams. Nowadays a sanctuary for past and present life, the sea remains a fantastic field of exploration for humankind. In many ways, the ever-lasting quest to understanding our environment as well as the vital need to develop synergies between human actions and the Ocean make the study of marine environment one of the most topical research effort of our time.

**An ocean of applications**

At the heart of marine research, oceanographic studies aim at understanding the Ocean in its physical and biological aspects. Perhaps the most immediate outcome of oceanography is the modelling of the climate regulating behaviour of the Ocean. In this context, there

is often a strong interest in obtaining a map of the sea-floor for environmental monitoring (Brown et al. [2011]) or improving long-distance navigation through bathymetric navigation (Nygren and Jansson [2004]).

Similarly to its equivalent on land, underwater archaeology provides an insight into human history by focusing on the study of archaeological remains such as shipwrecks, flooded constructions and ancient man-made objects. In this situation and *a-fortiori* when operating in deep water, there is a need for underwater vehicles to enable semi-autonomous or autonomous data gathering (Bingham et al. [2010]). In particular, the 3D reconstruction of the area of study from data gathered by an AUV provides archeologists valuable data for visual inspection and mission planning while the possibility to perform online semantic labelling improves time efficiency by enabling long, recovery-free missions.

For now more than a century, the ever-growing need for energy production has driven the development of offshore oil and gas rigs. Nowadays, recent advances in harvesting wind and tidal energy emphasize the importance of developing safe underwater robotics systems for deployment and maintenance of man-made infrastructures at the sea. Similarly, from the need for fast and long-distance telecommunication means arose the deployment of submarine telecommunication cables. Thanks to recent technical advances in the development of ROVs and AUVs (Whitcomb [2000]), structure deployment and maintenance on offshore fields are now carried increasingly autonomously.

Once a natural barrier keeping people apart, the Ocean has now become a space of intense traffic due to the development of marine technologies. As a result, considerable efforts have been put towards the development of reliable monitoring systems. In particular, military underwater tasks typically require the use of specific semi-autonomous systems for harbour surveillance (Madureira et al. [2009]), ship hull inspections (Hover et al. [2012]), surface vessels protection and hazards detection and identification such as underwater mines (Hagen et al. [2003]).

**A need for acoustic-based 3D semantic mapping**

In each of these fields, the progress of scientific and technological research is conditioned by the advances in marine robotics, providing increased autonomy, controllability and high-quality measurements. In this context, there is a strong interest in improving the observation and analysis capabilities of marine robotic platforms. In particular, acquiring an accurate representation of the 3D world and being able to obtain semantic information

on the environment from the scarce choice of underwater sensors are key challenges limiting the autonomy in marine robotics. Due to the poor propagation properties of electromagnetic waves in water, acoustic sensors remain the privileged sensing modality for short to long range sensing.

In this thesis, we propose to push the state-of-the-art of 3D semantic mapping in underwater environments by improving 3D sensing capabilities from standard 2D SONAR imaging and exploring object recognition in SONAR-based 3D reconstructions.

## 1.2 Thesis organisation

The organisation of this thesis is as follows:

- Chapter 2 provides an introduction to the reader on the principles of underwater acoustic sensing. Through the description and mathematical modelling of the multiple physical phenomena involved in SONAR sensing, we provide key elements to understand the characteristics of SONAR data. Based on these elements, we present the differences between the different types of SONAR systems and highlight the interest in using 2D imaging SONAR sensors, thereby justifying their choice in our research. Based on the characteristics of 2D imaging sensors, we present the basis of the 3D reconstruction problem by describing the aperture problem. Finally, we present our 2D imaging SONAR simulation framework with which our algorithms will be tested on a variety of realistic situations.

- Chapter 3 provides a review on the state-of-the-art in 3D reconstruction from SONAR data. Due to the specificities of each sensor, the review is split in 5 sections describing the variety of processing techniques investigated to obtain a 3D representation.

- Chapter 4 describes our first 3D reconstruction method from 2D SONAR data, referred to as space carving technique. Based on a simple observation, we describe the principle of the carving technique and its implementation. Experimental results on both simulated and real data are presented, along with an initial quantitative analysis.

- Chapter 5 introduces our second 3D reconstruction method from 2D SONAR data, referred to as a deconvolution technique. Based on the formulation of the 3D reconstruction problem as a spatially-variant blind deconvolution, a practical solution is proposed and formulated as a constrained optimization. A variety of experimental results is presented and compared to the space carving method both qualitatively and quantitatively. With an emphasis on field applications, the interest of both methods is then discussed as well as the influence of the sensor characteristics.

- Chapter 6 presents our research efforts in taking advantage of the possibilities of SONAR-based 3D reconstructions. In particular we investigate the extension of 3D mapping from SONAR data to large areas where navigation drift becomes consequent. We propose to extract semantic information from the reconstructed scene based on a set of rough CAD models, effectively providing the ability to operate 3D object recognition from SONAR sensing on a field. We explore multi-modal mapping with the addition of video data to the 3D reconstructions. Once again with an emphasis on field robotics, we provide an analysis on the applicability of our research work as well as possibilities of further developments based on our results.

## 1.3 Contributions

The main contributions of this thesis can be summarized as follows:

- **2D imaging SONAR simulator**: we developed a framework enabling realistic simulation of 2D imaging SONAR of any aperture. In comparison to previous SONAR simulation framework, our implementation takes advantage of a sensor-specific noise model providing realistic sensor-specific simulations in near real-time. The simulator has been successfully integrated in UWSim providing a complete simulation environment for the evaluation of algorithms and sensors. In particular, we show that the use of the simulation environment is of great interest as a pre-dive tool to optimize the result of underwater operations against operational parameters (sampling scheme, sensing modality, etc.) based on custom metrics of importance (reconstruction accuracy, surface coverage, time, power consumption, etc.).

- **Space carving 3D reconstruction method**: we developed a novel 3D reconstruction method based on the observation of empty spaces and an occlusion resolution processing. Importantly, the method enables online reconstruction and no strong constraint on the sampling pattern is required in the formulation. Additionally, no assumption is made on the scene, allowing reconstruction of scene with or without background surface. We demonstrated the interest of using this method on an AUV and the possibility to obtain 3D reconstructions at a centimetre level, similar to the accuracy of the sensor and local navigation drift.

- **Deconvolution reconstruction method**: inspired by SAS techniques, we propose a novel formulation of the reconstruction problem as a spatially-variant blind deconvolution. In order to address the notorious complexity of the problem, we adopt an approximate resolution through a constrained sparse solver and explore multi-pass deconvolution. We show the interest of our method on real data and exhibit the possibility to obtain detailed and high coverage reconstructions.

- **Quantitative analysis of 3D reconstruction from 2D imaging SONARs**: through a comparison to reference CAD models, we provide the first quantitative results on the accuracy and coverage of 3D reconstruction results obtained from 2D imaging SONAR data.

- **Study on the choice of an acoustic sensor for 3D reconstruction**: based on the simulation environment, we conducted a comparative study between pencil-beam and wide-beam SONARs highlighting the interest of using commonly embedded wide-beam sensors for high-coverage 3D reconstructions.

- **CAD-model-based underwater object recognition method**: we demonstrate the first CAD-model-based underwater object recognition method through a direct 3D CAD model mapping approach. Through our field experiment using SONAR data gathered by an AUV, we show the ability to identify the nature and 6D position of man-made structures based on a set of CAD models specifying the objects of interest.

- **Model-based video mapping**: using field data, we demonstrate the interest in performing model-based video mapping for visual inspection of 3D man-made objects.

# Chapter 2

# Sonar principles and simulation

> "If you cause your ship to stop, place the head of a long tube in the water and place the outer extremity in your ear, you will hear ships at a great distance from you"
>
> Leonardo Da Vinci, 1490

## 2.1 Introduction

With these famous words written in the $15^{\text{th}}$ century, Leonardo Da Vinci described the favourable propagation properties of sound waves in water. Since this early observation, successive technological improvements arose. The tragic sinking of the Titanic in April 1912 as well as the two successive World Wars rose awareness on the interest of exploiting sound waves to observe distant objects. As a result, many patents were filed and continuous improvements in maximum range and spatial resolution were achieved, leading to the development of so-called SOund Navigation And Ranging systems. Modern SONARs are now widely used in very different fields of application both in the civil and military domain. Their contained cost and size nowadays enables their integration on small underwater platforms.

We present here an introduction to SONAR principles, encompassing the various physical phenomena occurring during the emission and propagation of acoustic waves in a 3D space.

We then provide a generic measurement model for SONAR sensing before reviewing the differences between each type of SONAR sensor. Based on this review, we justify the choice of using 2D imaging SONARS for underwater environment observation. Adapting the general description of the physical phenomena to acoustic sensing from a generic transducer, we provide a mathematical model for a 2D imaging SONAR. We present the so-called aperture problem, highlighting the difficulty to estimate the 3D position of a scatterer from a single SONAR measurement. Finally we present a simulation framework enabling realistic simulation of 2D imaging SONARS of any vertical aperture. In comparison to existing SONAR simulation, our simulation framework enables the generation of 2D SONAR images at sufficient rate for enabling near real-time simulation and features a realistic sensor-specific noise model.

## 2.2 Principles and modelling

We present in this section the various physical phenomena that need to be taken into account when studying the formation of SONAR images. At each step, we provide mathematical models to be integrated later in a sensor-specific model.

### 2.2.1 SONAR equation

In 1967, Urick formulated the SONAR equation from an energetic point of view (Urick [1967]), modelling the relation between the emitted and received power levels in dB (relative to the reference intensity of a $1\mu$Pa plane wave):

$$SL - 2TL + TS - (NL - DI) = DT \tag{2.1}$$

with the parameters being:

- $SL$: Source Level (emitted power).
- $TL$: Transmission Loss (loss occurring during the propagation of the wave in open water).
- $TS$: Target Strength (target reflectivity or proportion of backscattered energy).

- $NL$: ambient Noise Level.
- $DI$: Directivity Index.
- $DT$: Detection Threshold or minimum SNR allowing the detection of a scatterer.

### 2.2.2 Source

In the case of an active SONAR, typical emission levels are around 200 dB. High emission levels are desired for both long-range sensing and achieving a good SNR (thanks to a high dynamic range). However, in practice, due to non-linear and cavitation effects (Urick [1967]), the use of high-power rectangular short pulses appears to be impossible in the water. Furthermore an ideal pulse (Dirac pulse) would feature an infinite bandwith which is in practice impossible to achieve due to the limited size of the transducer. As a consequence, the signals transmitted by acoustic transducers are band-limited. While the maximum range and the dynamic range of the SONAR are proportional to the transmitted energy, the range resolution is determined by its pulse duration (Callow [2003]). As illustrated in figure 2.1, the pulse width must be chosen small enough to enable the dissociation between scatterers observed consecutively along the direction of propagation. The duration of the pulse is naturally limited by the wavelength which is in turn defined by the frequency of the emitted signal. Small range resolutions can therefore only be achieved at high frequencies. On the other hand, the attenuation of sound is much higher at high-frequencies which limits the distance of observation (maximum range). When designing a acoustic source, there is therefore a trade-off between the desired range resolution and the maximum range. In order to maximum the resolution within practical bandwidths, techniques like *pulse-compression* have been developed (Cook [2012]), enabling range resolutions proportional to the signal bandwidth by emitting phase-modulated signal and operating a correlation between the received signal and the emitted signal. While the use of a wideband signal provides increased resolutions, in practise the bandwidth of the signal remains limited by the physical properties of the transducer (material and size) as well as the electronics. As detailed in Pailhas [2013], these aspects make the development of wideband systems considerably more complex. For these reasons, most SONARs operate at a single frequency and provide pulse-like signals through a simple amplitude modulation. Since we aim at performing centimeter-level 3D

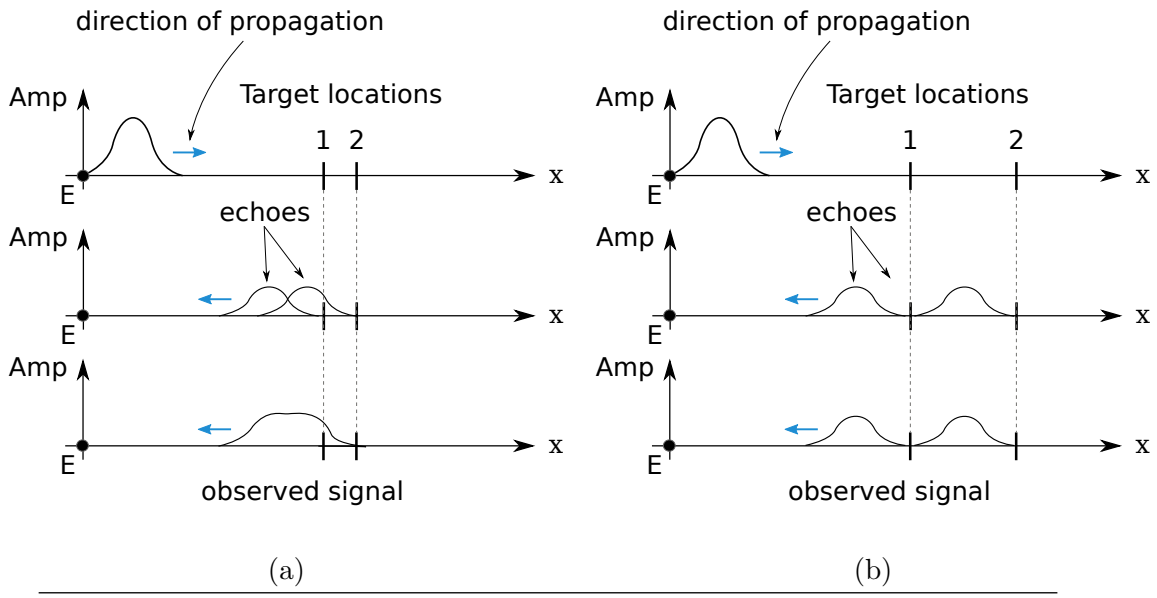



FIGURE 2.1: Illustration of the pulse width limiting the range resolution of a SONAR. A signal is emitted at E along the x-axis and backscattered by two targets. In the situation where the pulse width is greater than the distance between the two scatterers (a), the backscattered signals at each target sum up to a single large echo making the dissociation of the two scatterers impossible. On the contrary, when the pulse width is smaller than the distance between the two targets (b), the backscattered signal features two clear and distinct echoes.

reconstruction, the range resolution needs to be of centimeter level and we will therefore focus on high-frequency SONARs, emitting signals of at least a few hundreds kHz.

We therefore restrict our study to single and high-frequency designs and formulate the signal emitted by the transducer as:

$$p_e(t) = Amp(t)\ sin(t + \omega_0) \tag{2.2}$$

where $Amp(t)$ is a time-varying amplitude, often denoted as signal envelope and $\Omega_0$ is the phase at $t = 0$.

As described in section 2.2.7, real imaging systems are typically composed of an array of transducers and the signal observed at the receiver is typically obtained by auto-correlations over time periods to measure the energy backscattered at a given range. In this situation, phased arrays enable to increase the SNR through a so-called beamforming process (see section 2.2.10) and effectively reconstruct the emitted pulse. For this reason,

we formulate the model of our pulse as an ideal Dirac pulse:

$$p_e(t) = \delta(t - t_0) \tag{2.3}$$

with $t_0$ the emission time of the pulse.

### 2.2.3 Propagation model

Recalling the 3D wave equation describing the behaviour of a wave in an isotropic non-viscous fluid:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p = c^2 (\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2}) \tag{2.4}$$

where $p$ is a pressure field and $c$ is the spatial velocity of the wave front in the water.

It can be shown that functions of the following form in spherical coordinates notation are solution:

$$p(t, r) = \frac{1}{r} f(t \pm \frac{r}{c}) \tag{2.5}$$

This solution exhibits a typical spherical spread behaviour where the intensity decreases with the distance from the source. An important consequence of this is a quadratic decrease in power when ensonifying a distant target. In the case where the emission and reception are made at the same point, the two-way propagation therefore implies a power attenuation in $r^{-4}$.

In addition to the spherical spread, additional loss due to physical reactions in the water has been observed. As illustrated in figure 2.2, the absorption is frequency-dependent and mainly needs to be considered when operating at high frequency.

In order to compensate for transmission loss happening when observing distant targets, a TVG (Time Varying Gain) is often applied before the analog to digital conversion step, allowing the generation of images with range-independent intensities. For this reason, propagation loss terms are often dismissed in models and we will assume here that we are modelling sensors with ideal TVG, compensating for any propagation attenuation.

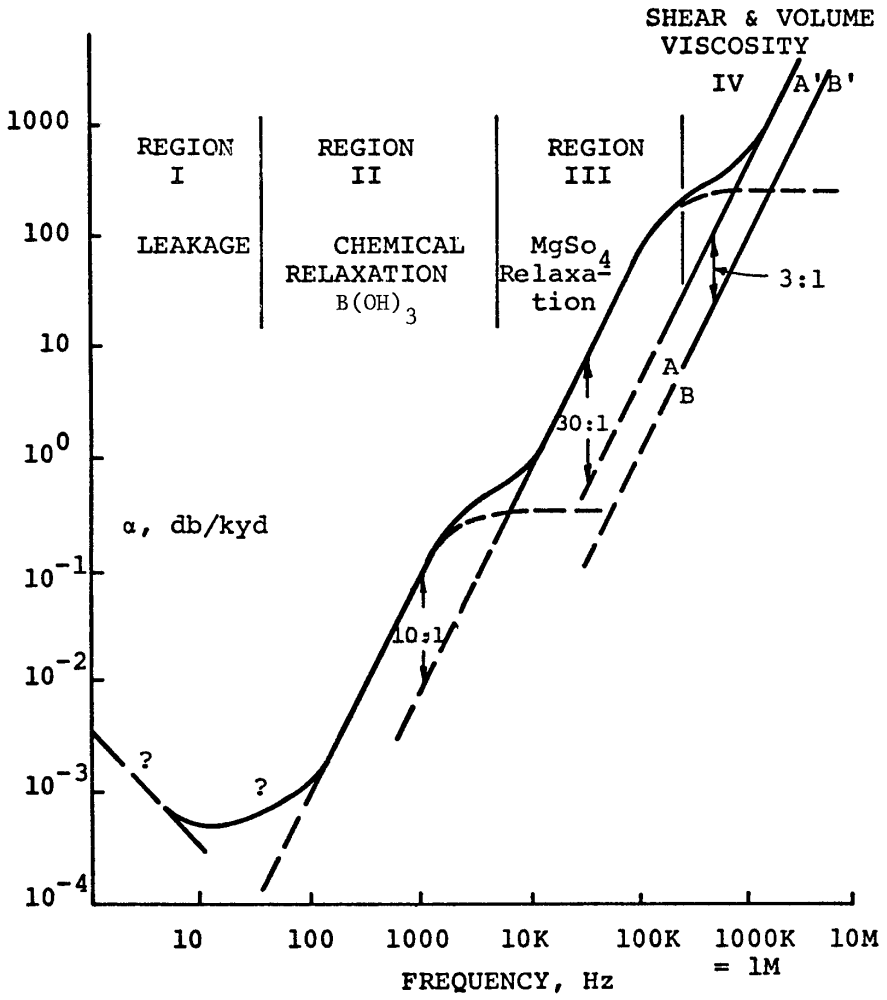



FIGURE 2.2: Regions of dominant processes of attenuation of sound in sea water - Urick [1967].

### 2.2.4 Sound scattering model

When meeting a solid surface, acoustic waves encounter a sudden change in mechanical impedance leading to the rejection of part of the incoming energy. Depending on the properties of the material (surface geometry, hardness), the wavelength considered and the incidence angle, the amount of energy bounced back by the scatterer varies widely. In general, the ratio between the size of the scatterer and the wavelength of the incoming wave determines the type of scattering. While Rayleigh scattering models are commonly used to model situations with large wavelengths compared to the scatterer size, geometric scattering is often assumed when considering large scatterers compared to the wavelength. In this situation, the laws of geometric optics are sufficient to describe the phenomenon.

This phenomenon, generally denoted as acoustic reflectivity in the literature has been widely studied underwater.

In Urick [1954], Urick presented the backscattered energy as a function of pulse length, frequency and grazing angle. His early experiments suggested that the surface roughness determined the characteristics of the backscattering and exhibited a diffuse reflection, rather than a specular diffusion. Further experiments ([McKinney and Anderson, 1964, Stanic et al., 1988]) made on different types of seabed at a large range of grazing angles confirmed the results of Urick. Alongside with Marsh and Patterson models, the Lambertian model (see equation 2.6) is used in Gott et al. [1993] where a parameter estimation is made from the backscattered data of a 12kHz sonar array.

$$I_r = \mu I_i sin(\theta) sin(\phi) dS \tag{2.6}$$

where $0 \leq \mu \leq 1$ is a reflectivity coefficient modelling the proportion of energy backscattered by a material, $I_i$ is the intensity of the incident wave, $\theta$ and $\phi$ are the two grazing angles (complementary of incidence angle) characterising the direction of the incident wave w.r.t the local surface patch $dS$. Due to its simplicity, the Lambertian assumption is commonly assumed in underwater imaging (Aykin and Negahdaripour [2013]).

When operating at low grazing angles, the authors of Trevorrow [2004] showed that a Rayleigh model was better suited than Lambert's law by studying the skewness of the backscattered distribution. Similar recent studies such as Lyons and Abraham [1999] found that log-normal distributions and Rayleigh mixture model provide a better fit to the tail of statistical distribution of backscattered signals than Lambert's model. However in general, the roughness of the surface compared to the wavelength remains the dominant factor to be considered in the choice of a model. Altough using high-frequency SONARs with wavelengths of a few millimeters, we consider in this study the observation of surfaces with very low roughness such as clean man-made structures. This type of surface therefore exhibits a specular behaviour that justifies the Lambertian assumption. We define a scatterer as being an element of large size in comparison to the wavelength of the sensor, therefore when considering a 900kHz sensor a scatterer would be typically half a centimeter at least.

### 2.2.5 Acoustic shadowing

One important consequences of sound scattering and propagation in the water is the shadowing phenomenon. Due to the backscattering of sound waves by the surface of the objects, no acoustic sensing is possible in the inner part of the objects. In general, the presence of obstacles (ranging from small particles to solid objects) affects the propagation of the wave by reflecting a fraction or all the incoming energy, making further acoustic sensing in the direction of propagation impossible. Therefore when the emitter and receiver are coincident, only the surface visible from the location of the sensor can be observed, leaving the rest of the scene unknown.

As illustrated in figure 2.3, in presence of a scatterer within the beam boundaries, the region behind the object does not reflect any energy and therefore remains unobserved. However, on the side of the scatterer, the wave keeps propagating, enabling the observation of further objects in these directions.

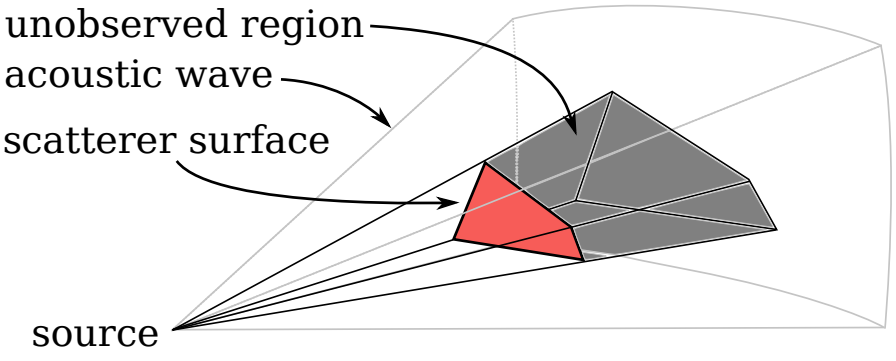



FIGURE 2.3: Acoustic shadowing. A scatterer is present on the propagation path of an acoustic wave creating an acoustic shadow. The region behind the solid surface can therefore not be observed from the source position.

When adopting uniform representations such as 2D images, this absence of backscattered energy leads visible shadows behind the obstacles. An illustration of this phenomenon is shown in figure 2.4 where a boat hull present on the seabed is imaged with a BlueView P900-130 SONAR. Due to the occlusion of the boat in the field of view of the SONAR, a shadow is visible in the area behind it.

As a consequence, acoustic shadows must be taken into account when interpreting SONAR measurements. The observation of an absence of acoustic returns can be caused by both open water propagation and shadowing effect. Therefore and under the sensor

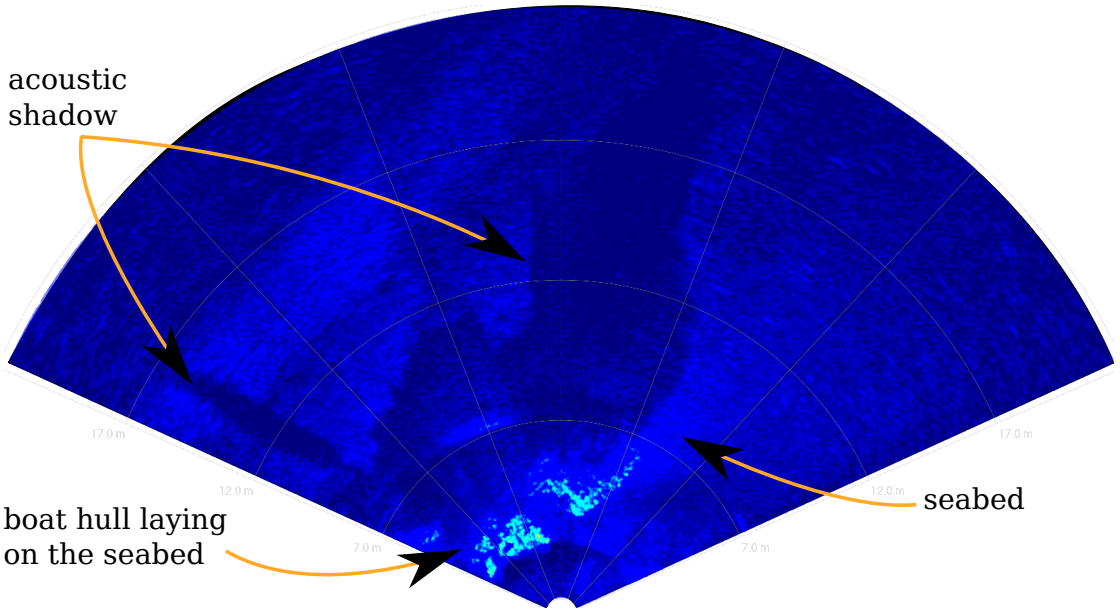



FIGURE 2.4: Illustration of acoustic shadowing in real data. BlueView P900-130 image during the inspection of a sunk boat hull, laying on the seabed. The image exhibits large shadows area behind the boat due to the occlusion of the seabed by the boat hull.

spatial resolution, it is *a-priori* impossible to determine directly whether or not a region with no acoustic return is occupied or not.

### 2.2.6 Multipath propagation

Another consequence of sound bouncing on surfaces is the multipath propagation phenomenon. As showed in figure 2.5, when multiple scatterers are present in the scene, the emitted acoustic wave can successively bounce from one scatterer to another. In this situation, it is possible to observe an acoustic return of a travel time equal to the sum of all the travel durations between the different elements. As a consequence the wave reflected by the last scatterer will appear as coming from further than the direct distance to the transducer.

When unable to resolve the direction of arrival of the observed wave (for example under the angular resolution of the sensor), this phenomenon typically leads to wrong interpretation of the delayed return. In this situation, an acoustic return at a further distance in the direction of the sensor will be represented. For this reason, this type of return is sometimes denoted as a *ghost* return.

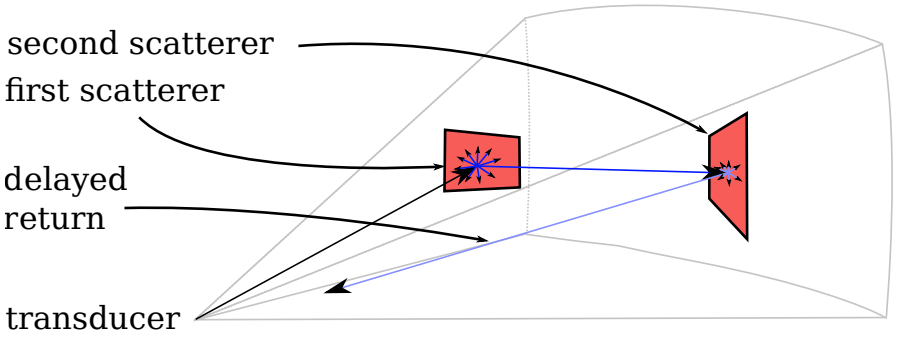



FIGURE 2.5: Multipath propagation. An emitted acoustic wave is reflected by two scatterers in the scene before returning to the transducer location. In this situation, a delayed return is observed as featuring a travel time equal to the sum of the three point to point propagation paths (transducer to scatterer 1, scatterer 1 to scatterer 2 and scatterer 2 to transducer).

This phenomenon is typically observed in situation where multiple scatterers are present in the scene such as when observing complex geometrical shapes or operating in closed environments (the presence of interfaces on the sides such as walls, seabed or water surface will induce acoustic reflections). As an illustration of this phenomenon, figure 2.6 shows two SONAR images in polar coordinates acquired in a small water tank in Heriot-Watt University. Two different shapes are represented, figure 2.6-a shows an aluminium sphere and its security rope while the figure 2.6-b exhibits a more complex object (Hyball ROV). Both objects have been placed on the bottom of the tank, 1 metre away from the side walls. In addition to the objects and the bottom of the tank, acoustic returns reflected by the water surface, the tank side walls as well as multiple parts of the vehicles.

Although weaker by nature (multiple attenuations), these returns severely pollute the image and make its interpretation complex. In a situation of multipath observation, ghost returns are visible but a similar image could be obtained with a real scatterer in place of the ghost return. It is therefore *a-priori* impossible to determine the position of the multiple scatterers from a single observation.

### 2.2.7 Measurement model

Once the pulse transmitted, any incoming wave located within the physical aperture boundaries of the transducer is observed. As elaborated in antena theory (Kino [1987]),

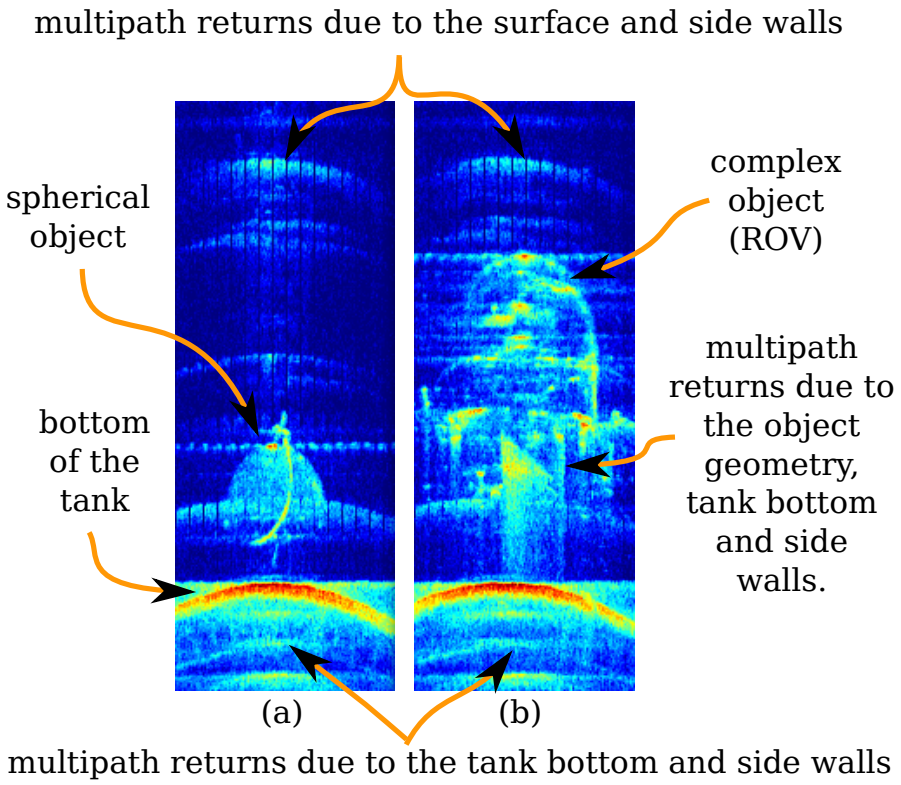



FIGURE 2.6: Illustration of a multipath propagation on real data. Two objects are being imaged in a small water tank. A simple, spherical shape (a) and a more complex geometrical shape (b) are placed at the bottom of a small water tank. In both cases, reflections coming from multipath between the tank side walls and respectively the water surface and the bottom of the tank are observed. In (b), multipath due to multiple reflections between the parts of the object (Hyball ROV) are visible.

the plane wave assumption is only valid in the far field of the sensor which starts at 1m for a SONAR operating at 900kHz. In this study we therefore assume far-field conditions and the observed waves as being plane waves. The transducer then essentially records the acoustic state in its vicinity which results from the sum of all these waves. The physical aperture of the transducer induces an attenuation pattern varying on the angle of observation.

As illustrated in equation 2.7, mathematically, the measured intensity is therefore a 2D integral of the acoustic potential in its vicinity:

$$I_{Transducer}(t) = \int_{A_v} \int_{A_h} B_v(\theta) B_h(\phi) V(t, \theta, \phi) \mathrm{d}\theta \mathrm{d}\phi \tag{2.7}$$

where $A_v$ and $A_h$ are respectively the physical vertical and horizontal apertures of the transducer, $B_v$ and $B_h$ their associated beam patterns and $V(\theta, \phi, t)$ is the acoustic

potential observed at an elevation angle $\theta$, an azimuth angle $\phi$ and a time $t$.

### 2.2.8 Directivity model

In order to obtain an increased resolution and better power efficiency, acoustic transducers often feature radiation patterns with a high directivity (narrow beam pattern), effectively focusing the emitted energy in a specific direction. This angular attenuation affects the signal both during the transmission and reception phase. When unknown, the beam pattern can be estimated such as in Lanzoni and Weber [2010] where the transmission and reception beam pattern of a Reason 7125 sonar have been measured. In many cases, sonar sensors are effectively composed of an array of sensors. As mentioned in Thorner [1990], transducers beam pattern often exhibit side-lobes resulting in sensing signals coming from neighbours transducers or multi-paths waves. In practice, manufacturers aim at reducing the side-lobes as much as possible while approaching a step profile in the Fourier domain (ideal low-pass filter). For this reason, the beam pattern is typically modelled by a Bessel function of the first kind:

$$B(\theta) = \left| \frac{sin(\frac{\kappa.\theta}{\theta_{3dB}})}{\frac{\kappa.\theta}{\theta_{3dB}}} \right| \tag{2.8}$$

with $\kappa \approx 1.389$ and $\theta_{3dB}$ being the angle of mid-power attenuation such that $[B(\theta_{3dB})]^2 = 0.5$, commonly used to define the aperture of the sensor. Figure 2.7 provides an illustration of the beam pattern with $\theta_{3dB} = 5°$exhibiting small sides lobes, at least ten times the amplitude of the main lobe. As a consequence, we will consider this model when evaluating the angular attenuation.

### 2.2.9 Phased arrays

In order to increase both directivity and footprint, SONARs are usually composed of an array of transducers. They effectively produce 2D images from the measurements of each element on a unidimensional array (Belcher et al. [2002]) or 3D range images with a 2D array (Davis and Lugsdin [2005]). So as to maximize the field of view, each transducer can be used alternatively as an emitter or a receptor. Depending on the characteristics of the sensor, specific emission patterns are designed by the SONAR manufacturers in order to allow the acquisition of the data on each beam independently and minimize

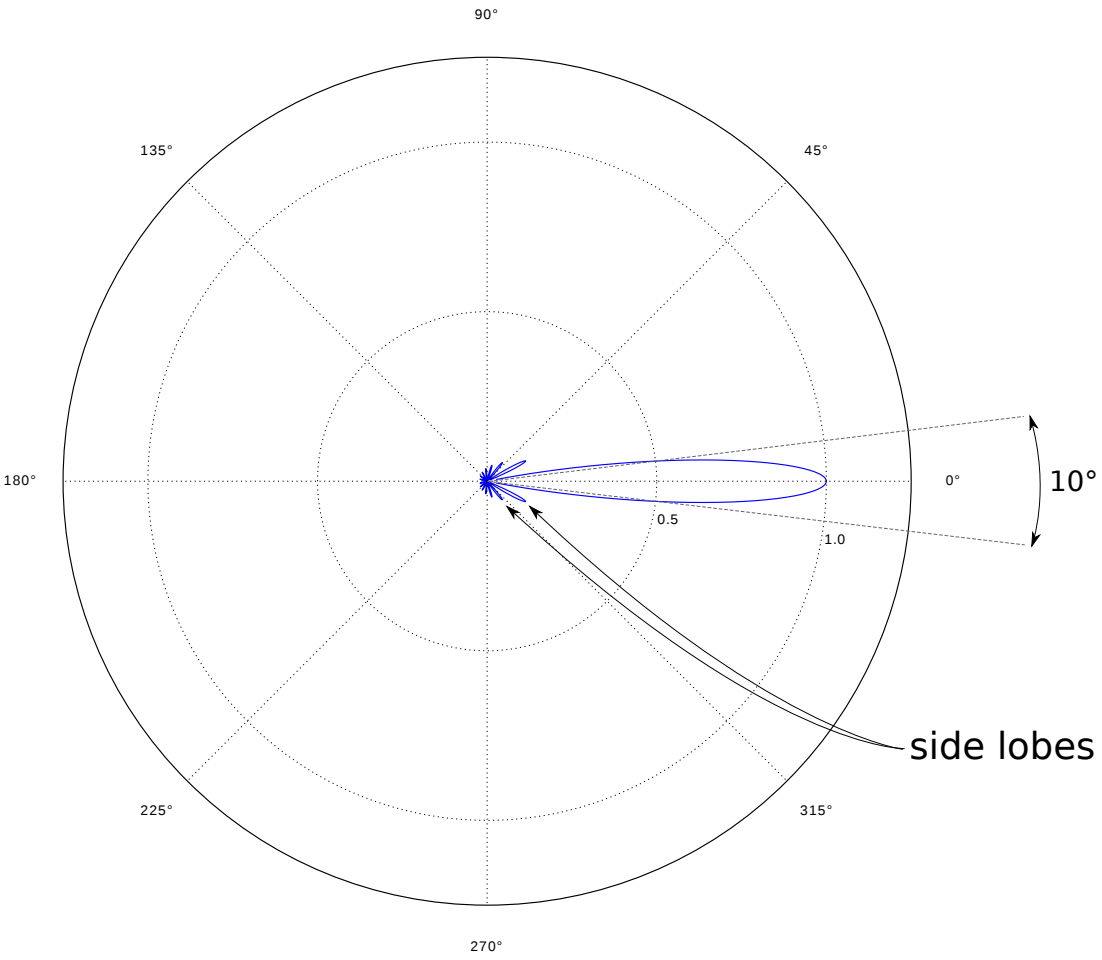



FIGURE 2.7: Typical beam pattern of a 10°aperture transducer pointing along the horizontal axis (0°direction) and displayed in polar coordinates with normalized amplitude.

crosstalk artefacts. Although all beams are acquired independently and sequentially, the high frequency SONARs considered in this study operate short range sensing (1 to 50m). Due the constraint of having a high range resolution, our 3D reconstruction applications target inspections at typically 1 to 10m range. At these distances and using high-frequency sensors, SONARs typically feature updates rates of a few tenths of Hz. In these conditions, the change in the scene between the acquisition of two beams (either due to the vehicle motion or to the evolution of the scene) can be considered negligible. We will therefore ignore the impact of the emission pattern of the sensor in this study.

### 2.2.10 Beamforming

Exception made of acoustic lens-based designs such as Belcher et al. [1999], traditional SONARs use transducers featuring a non-negligible aperture of at least 1°. The angular uncertainty due to the aperture of the transducers is then reduced by using a so-called beamforming technique which consists in combining the observations acquired by different transducers along the array. As illustrated in figure 2.8, the angle of arrival $\theta$ of the wavefront induces a delay between the observations of the wavefront at each transducer. Based on this principle, the angle $\theta$ can be estimated through a so-called electronic

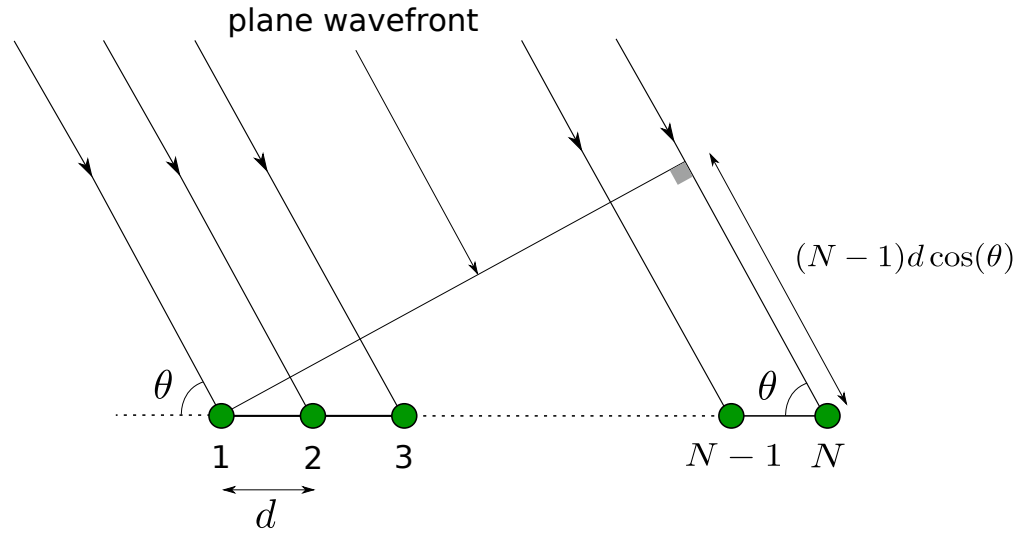



FIGURE 2.8: Estimation of the direction of arrival using a phased array. Q plane wave is observed at each transducer $(1...N)$ with a delay depending on the spacing between each transducer $(d = \frac{\lambda}{2})$ and the direction of arrival.

steering (or scanning) where the signals of each transducers are combined with a range of delays $(N-1)d\cos(\theta) = (N-1)\frac{\lambda}{2}\cos(\theta)$ corresponding to the direction of measurement. Thus the contributions of a wave coming from an angle $\theta$ will remain coherent and add up whereas observations in directions where no wave is coming from will appear incoherent and result as noise. This processing therefore increases the Signal to Noise Ratio (SNR) and provides a higher resolution along the direction of the array. This phased array processing is therefore called beamforming for its ability to estimate the missing direction of arrival making it equivalent to a sensor with a small aperture along the direction of the array.

Based on this principle of phased array, many beamforming techniques have been investigated. As described in Thorner [1990], the beamforming scheme can be either adaptive or fixed. In adaptive beamforming, the received signal is used to refine the characteristics of the array processing, allowing dynamically optimized measurements for various purposes (interference rejection, SNR maximization, etc.) whereas fixed beamforming makes use of constant time/delay weights to form the image. Since the first publication of electronic beamforming of acoustic sources (Billingsley and Kinns [1976]), three types of beamforming methods have emerged. Simple delay-and-sum approaches estimate the reflected wave intensity at a given range and bearing angle by summing the measurements of each transducers weighted by a delay term. The energy coming from this range and bearing angle will then add-up coherently or incoherently depending on whether or not a scatterer was present at this location. While simple, this technique a-priori requires considering every possible angle on a sufficiently high resolution sampling grid, resulting in heavy computations. In opposition to time domain approaches, frequency-based approaches operate the beamforming by processing the dual frequency representation of the measured signal. Importantly and unlike time-domain beamformers, the beam resolution achieved by frequency-based methods is not limited by sampling period. Improvements to reduce the sensitivity to noise have been made such as the MUSIC beamformer (Schmidt [1986]) where the frequency representation of the signal is linearly decomposed, identifying the noise characteristics and improving the angular estimation. The last beamforming approach are called maximum likelihood approach where an estimation of the spatial covariance matrix is obtained by minimization towards a model. In practice, the beamforming operation is in general computationally intensive and often implemented by the manufacturer on dedicated hardware platforms such as FPGAs (Graham and Nelson [1998]). Implementation details are at the discretion of the manufacturer and depend on the intrinsic design of the sensor. For this reason, all manufacturers provide the beamformed data by default.

While the phase information contained in the signals pre-beamforming is of interest for the problem of 3D reconstruction in general, it is in practice difficult to leverage due to the constraints imposed by coherent processing (position known to a fraction of the wavelength accuracy) as well described in SAS work Marston and Kennedy [2016], Sæbø et al. [2013]. Since we chose to operate with high-frequency SONARs of at least a few hundreds of kHz, the wavelenghts considered amount to a fraction of a centimeter making

coherent processing impossible with current navigation systems. For this reason, we choose to assume our data as beamformed and our modelling of a 2D sonar will therefore assume direct bearing sensing, resulting in the following formulation:

$$I_{Sonar}(t,\phi) = \int_{A_v} B_v(\theta) V(t,\theta,\phi) \mathrm{d}\theta \tag{2.9}$$

where $A_v$ is the vertical aperture of the array of transducers, $B_v$ the associated beam pattern and $V(\theta,\phi,t)$ is the acoustic potential observed at an elevation angle $\theta$, an azimuth angle $\phi$ and a time $t$.

### 2.2.11 SONAR noise model

As studied in Wenz [1962], acoustic ambient noise levels in the oceans depend on the local conditions such as the presence of bubbles, surface agitation and turbulent pressure fluctuations. In addition to the ambient noise and depending on the operated frequency band and the quality of the SONAR electronics, various types of noise can corrupt the data. In spite of the use of beamforming techniques, increased sensor-specific noise patterns can be observed in the SONAR images. In particular, the measurements obtained in areas of high attenuation (side of the transducers beam patterns) exhibit lower intensities resulting in higher noise levels. An illustration of this phenomenon is showed in figure 2.9 where a sonar image has been measured in open-water in Loch Eil, Scotland.

In addition to sensor-specific noise patterns, speckle noise is frequently observed in sonar images as described in Abbott and Thurstone [1979]. From the use of simple image processing techniques (averaging multiple frames, median filter) to more complex filters ([Huang et al., 2009, Isar et al., 2005]), speckle noise can be mitigated in different ways, often at the cost of an edge smoothing. In general, the observed noise is the result of the combination of multiple kinds of noise in the acquisition chain. For this reason, a Gaussian noise is commonly observed in the final data and is assumed to be decorrelated from the properties of the scene. Although the presence of a scatterer in the field of view of the sensor could potentially modify the noise distribution, the study of this phenomenon would require a series of experiments in an open-water environment and an elementary scatterer to be placed in various locations in front of the sensor. It appears difficult to carry such an experiment and the dependency on the location of a scatterer

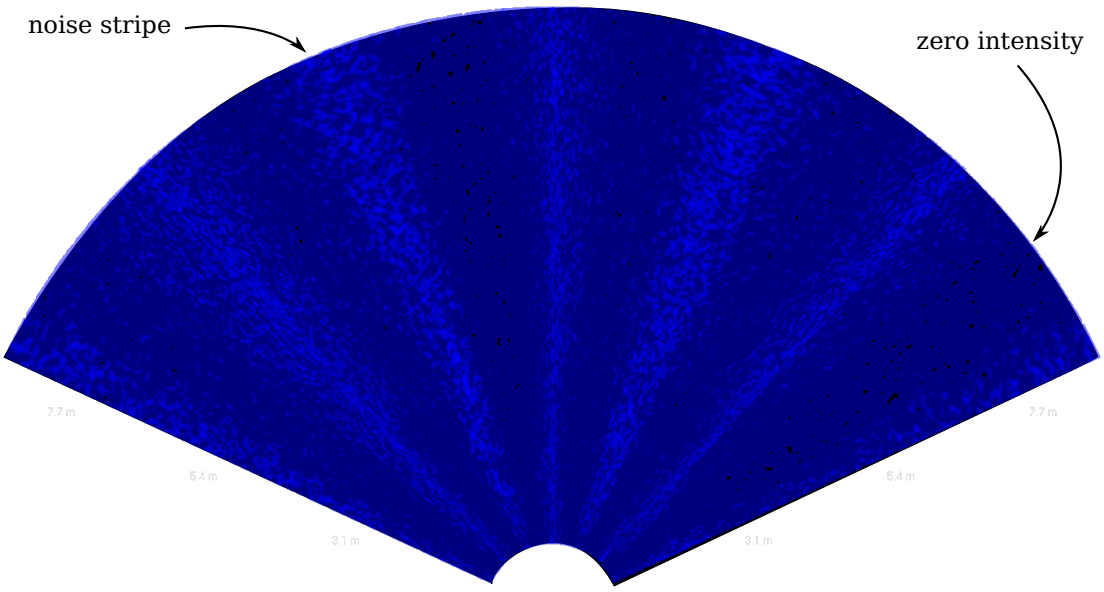



FIGURE 2.9: BlueView P900-130 noise pattern. In absence of any scatterer in the scene, the sonar image exhibits 7 noise stripes.

in the scene would make the denoising process a complex treatment. Our chosen noise model will therefore be sensor-specific and modelled based on the observation of noise statistics in a scatterer-free environment.

## 2.3 Overview of the different types of SONARs

In order to exhibit the specificities and usage of each type of sensor, we provide here a brief overview of the various types of SONARs existing along with their main application fields. Following this review, we highlight the particular interest in using 2D imaging SONARs for their frequent usage in a wide variety of applications.

### 2.3.1 Review

**Single-beam echosounders** As illustrated in figure 2.10, single-beam echosounders provide direct range readings in a single direction.

Consisting of a single transducer, single-beam echosounders often equip boats to provide a rough estimate of the draught. These sensors are inexpensive and have only been scarcely used for research work to provide low-resolution bathymetry measurements,

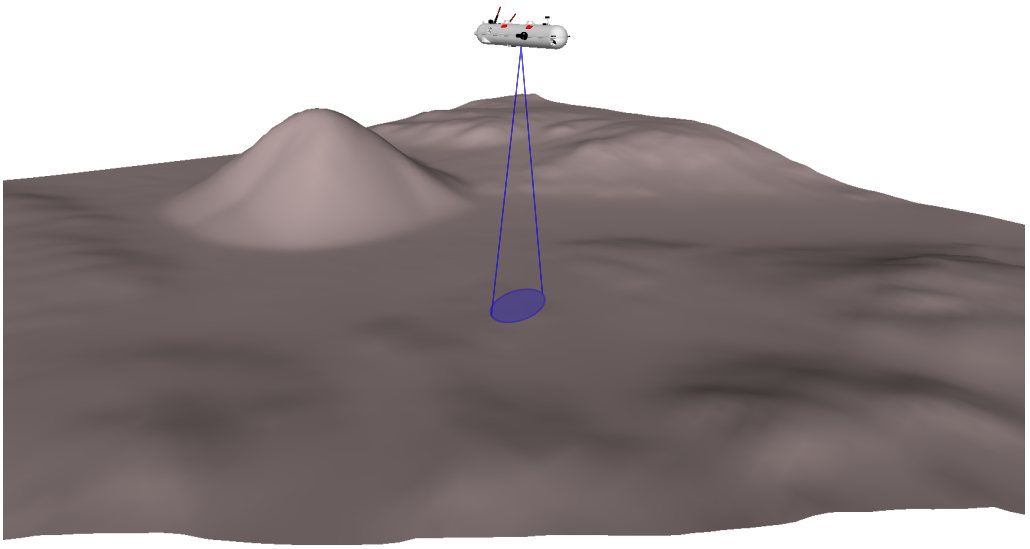


FIGURE 2.10: Illustration of the footprint of a single-beam sensor mounted on an AUV in a downward configuration.

study seabed composition (Amiri-Simkooei et al. [2011]) or observe local faun (Hutin et al. [2005]).

**Side-scan SONARs** Side-scan SONARs aim at providing backscatter data of a very narrow stripe from a single ping. The narrow aperture (typically a fraction of degree to 1°) is obtained by beamforming of an array of receiving transducers. Due to these specificities, side-scan SONARs can only be employed in a side-way-looking configuration as visible on figure 2.11.

For this type of sensor, low frequency designs are prevalent, allowing typical sensing ranges of a few hundreds meters. Side-scan SONARs are traditionally inexpensive and employed to image large areas of seabed for inspection, detection and classification of objects lying on the seabed. Therefore side-scan SONARs are frequently integrated on cost-efficient platforms with the aim of delivering long-range seabed imaging as in Collier and Humber [2007] for coral reef monitoring, pipeline tracking as in [Bagnitsky et al., 2011, Petillot et al., 2002] or mine detection as demonstrated by Reed et al. [2003].

**2D Multibeam imaging SONARs** In order to provide direct 2D readings, multiple transducers can be arranged as an unidimensional array. As visible on figure 2.12, this type of sensor naturally provides larger footprints than single-beam sensors, allowing direct range sensing at multiple bearing angles. Beamforming techniques then enable the

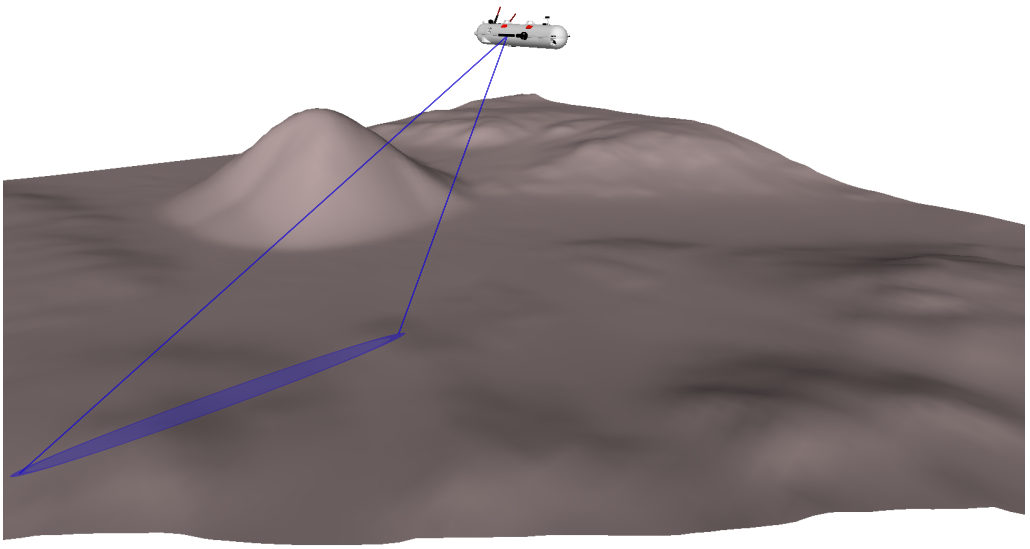


FIGURE 2.11: Illustration of the footprint of a side-scan sensor mounted on the side of an AUV.

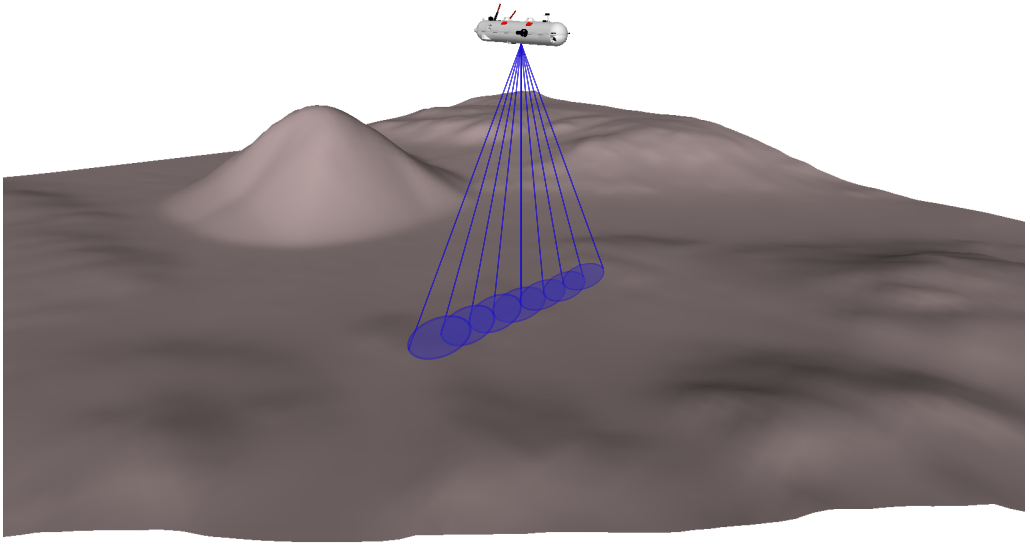


FIGURE 2.12: Illustration of the footprint of a multibeam sensor mounted on an AUV in a downward configuration.

generation of range and bearing 2D images. Depending on the design of the sensor, the bearing angle resolution typically ranges from a fraction of a degree to 2°. The range resolution depends on the frequency operated and typically ranges from a few millimetres to a few centimetres. Multibeam SONARs are then divided into two categories, based on their vertical aperture. Pencil-beam sensors offer a low footprint by featuring typical apertures around 1° whereas wide-aperture SONARs ensonify large volumes of water at each ping with typical vertical apertures of 7° to 20°.

Thanks to their low elevation angle uncertainty, pencil-beam sensors are traditionally used to acquire bathymetry ([Grasmueck et al., 2006, Vaneck et al., 1996]) at ranges up to a hundred metres, to perform short to mid-range mapping (Papadopoulos et al. [2011]), to study ecosystems or assist fishing activities (Gerlotto et al. [2000]).

On the other hand, wide-aperture SONARs are more commonly employed for collision avoidance tasks (Petillot et al. [2001]), sea floor imaging and mosaicking (Hurtós et al. [2014]) or online target tracking (Folkesson et al. [2007]).

**3D SONARs** In order to provide direct 3D information, a 2D set of acoustic transducers can be integrated in a single sensor. As illustrated in figure 2.13, these sensors offer direct range readings along two axis, effectively providing direct 3D information. These

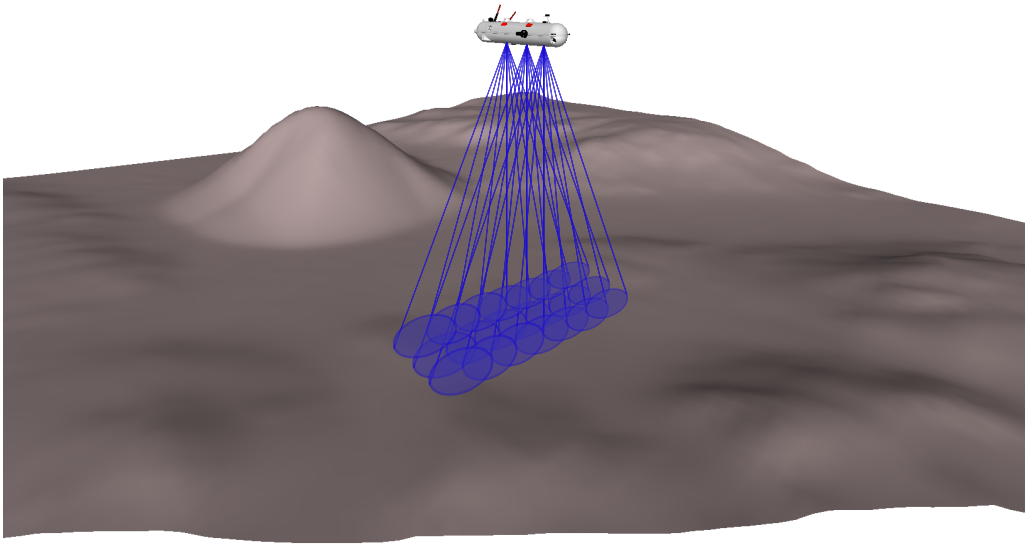


FIGURE 2.13: Illustration of the footprint of a 3D SONAR mounted on an AUV in a downward configuration.

| | Weaknesses | Strenghts |
|---|---|---|
| Single-beam | - low resolution<br>(20cm × 10°) | - low cost<br>- very long range (up to 7000m) |
| Side-scan | - medium to high resolution depending on range<br>(3 to 10cm × 1°)<br>- only provides 1D data and in slanted range | - low cost<br>- long range (up to 500m)<br>- higher resolution (SAS) provided accurate navigation is available<br>- compact and commonly embedded on AUVs |
| 2D multibeam | - short range (up to 100m)<br>- high cost | - high resolution<br>(0.5 to 3cm × 0.6°to 2°)<br>- relatively small and easy to integrate on AUVs<br>- provides 2D information of direct use for live monitoring |
| 3D array | - very short range (up to 100m but typically used up to 10m to yield high resolutions)<br>- very high cost<br>- bulky | - high resolution (3cm × 0.5°to 20°)<br>- provides direct 2.5D information of interest for collision avoidance |

TABLE 2.1: Strengths and weaknesses of the different types of SONAR sensors

3D SONARs usually operate at high frequency to provide higher resolutions (centimetre level) in a contained size. 3D SONARs have been employed to perform tasks where instant 3D sensing is desired, in particular to observe phenomena in real-time such as in Soloviev et al. [2012] where ship wakes were studied with an Echoscope SONAR (Davis and Lugsdin [2005]). Other applications include real-time positioning (Woodward et al. [2010]) or fast 3D mosaicking (Hansen et al. [2005]). Although various 3D SONARs are now commercially available, their increased complexity in comparison to standard 2D SONARs makes these sensors expensive and bulky. Their use on AUVs is therefore in general impractical if not prohibitive on cost-efficient platforms.

### 2.3.2 Interest in using 2D imaging SONARs

In order to identify the pros and cons of each type of SONAR, we present in table 2.1 a summary of the strengths and weaknesses of each configuration.

Although inexpensive, side-scan SONARs only provide 1D information about a narrow stripe at each ping and the recovery of 3D information from side-scan data is inherently ill-posed. On the other hand, direct 3D sensors such as the Echoscope provide real-time 3D sensing capabilities but remain prohibitively expensive and too large for being

integrated on AUVs. Multibeam sensors bridge the gap between these two categories of sensors by offering small to large footprints, 2D centimetre-level resolutions in a cost and size that makes them suitable for employment on a wide variety of surface and underwater vehicles. SONAR sensors being expensive in general, there is a lot of interest in using sensors commonly embedded on underwater platforms rather than adding a specific sensor for each new application. For these reasons, our work focuses on the use of standard 2D imaging SONARs for 3D reconstruction. For reference, appendix A - section 6.7 provides typical specifications of a 2D imaging SONAR.

## 2.4 Description of 2D imaging SONARs

Following the justification of the interest in studying the use of 2D imaging SONARs for 3D reconstruction purposes, we present here an imaging model specific to this sensor. In particular, we combine the generic acoustic sensing models presented in sections 2.2.2 to 2.2.11 with the specificities of 2D imaging sensors to present a mathematical model suitable to the description of the generation of 2D high-frequency SONAR measurements.

Based on this model, we present the aperture problem inherent to the imaging process and illustrate it to exhibit the difficulty of performing 3D reconstruction from 2D SONAR data.

### 2.4.1 2D imaging SONAR model

A 2D imaging SONAR is essentially a rectilinear array of transducers. These transducers are triggered at different times in order to provide readings at multiple bearing angles with an increased resolution when using beamforming techniques. As visible in figure 2.14-a, imaging sensors deliver 2D images where each pixel represents a range and bearing angle small interval. As depicted in figure 2.14-b, the pixels at each range range and bearing angle represent backscattered intensities coming from all elevation angles along the vertical aperture $A_v$.

Based on the models established in sections 2.2.3 to 2.2.8, we present here the combined imaging model of a multibeam SONAR in absence of occlusions and multipath propagation. Recalling the measurement model presented in equation 2.9 and considering an

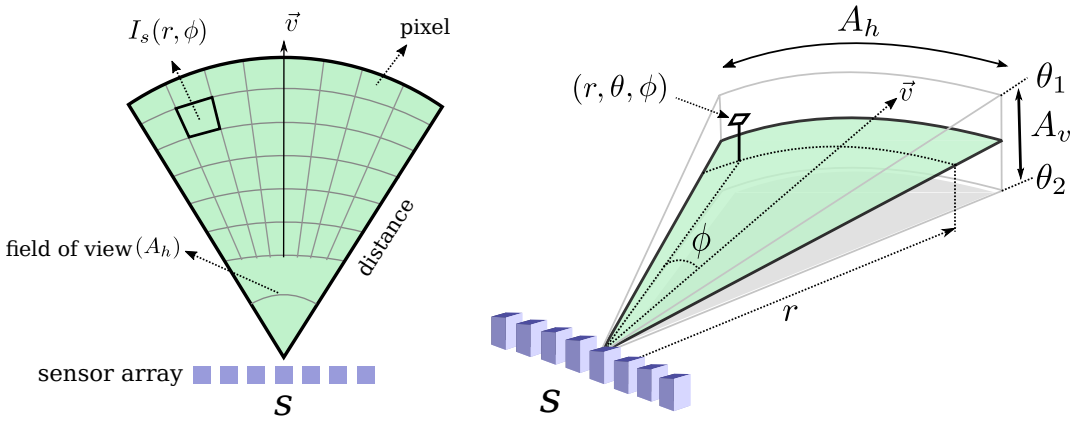



FIGURE 2.14: 2D Imaging SONAR geometry.

isotropic propagation of the emitted pulse $p_e(t)$ in the medium, the received signal at time t and bearing angle $\phi$ can be written as:

$$I_{Sonar}(t,\phi) = \int_{A_v} B_v(\theta)^2 \int_0^{\infty} R(r,\theta,\phi)\, \frac{1}{2r}\, p_e(t - \frac{2\,r}{c})\, \mathrm{d}r\, \mathrm{d}\theta \tag{2.10}$$

with $R(r,\theta,\phi)$ the scatterer reflectivity at the 3D point $(r,\theta,\phi)$. The signal observed by the transducer at time $t$ is therefore a sum of the acoustic returns coming from distant scatterers present in the scene with a delay $(t - \frac{2\,r}{c})$ associated to the two-way propagation and an attenuation in $r^{-1}$.

As explained in section 2.2.3, assuming compensation of the propagation loss (as obtained when using a TVG), the attenuation term can be omitted. Furthermore, we consider here an ideal dirac pulse $p_e(t - \frac{2\,r}{c}) = I_0\, \delta(t - \frac{2\,r}{c})$ which results in the following formulation:

$$I_{Sonar}(t,\phi) = \int_{A_v} B_v(\theta)^2 \int_0^{\infty} R(r,\theta,\phi)\, I_0\, \delta(t - \frac{2\,r}{c})\, \mathrm{d}r\, \mathrm{d}\theta \tag{2.11}$$

Based on this model, readings at range $r$ can therefore be obtained at $t = \frac{2\,r}{c} + t_0$ where $t > t_0$ and $t_0$ is the emission time of the pulse:

$$I_{Sonar}(r,\phi) = I_0 \int_{A_v} B_v(\theta)^2\, R(r,\theta,\phi)\, \mathrm{d}\theta \tag{2.12}$$

One can then describe the reflectivity $R$ of the scatterers as following a Lambertian diffusion model:

$$I_{Sonar}(r,\phi) = I_0 \int_{A_v} B_v(\theta)^2 \, S_s(r,\theta,\phi) \, \mu(r,\theta,\phi) \, \frac{\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}}{\|\vec{v}_{\theta\phi}\| \, \|\vec{n}_{r\theta\phi}\|} \, \mathrm{d}\theta \tag{2.13}$$

where $S_s(r,\theta,\phi)$ represents an indicator function that is unity if there is a surface at location $(r,\theta,\phi)$ (relatively to the position of the source) and zero otherwise. The term modelling the albedo of the surface (proportion of reflected energy) is defined with respect to the source location s by the reflectivity coefficient associated to the material of the scatterer $\mu(r,\theta,\phi)$ and the dot product $\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}$ where $\vec{v}_{\theta\phi}$ then represents the direction of propagation of the acoustic beam at the angles $\theta$ and $\phi$ and $\vec{n}_{r\theta\phi}$ is the surface normal of the scatterer at $(r,\theta,\phi)$.

### 2.4.2 The aperture problem

As described in the previous sections, an imaging SONAR delivers intensities by summing the multiple acoustic returns coming from different elevation angles. The width of the vertical aperture therefore has a direct impact on the ability to locate the scatterer position within the vertical aperture. This phenomenon is commonly called the aperture problem.

When considering large apertures, this blurring process induces an ambiguity when determining the position of the scatterer along the vertical aperture of the sensor. As illustrated in figure 2.15, two identical scatterers placed at opposite elevation angles relatively to the SONAR image plane result in the same measurement.

This example exhibits the 3D to 2D compression resulting from the SONAR imaging process. It demonstrates the inherent illness of the 3D reconstruction problem and the *a-priori* impossibility to determine the position of the scatterer based on a single image without additional knowledge.

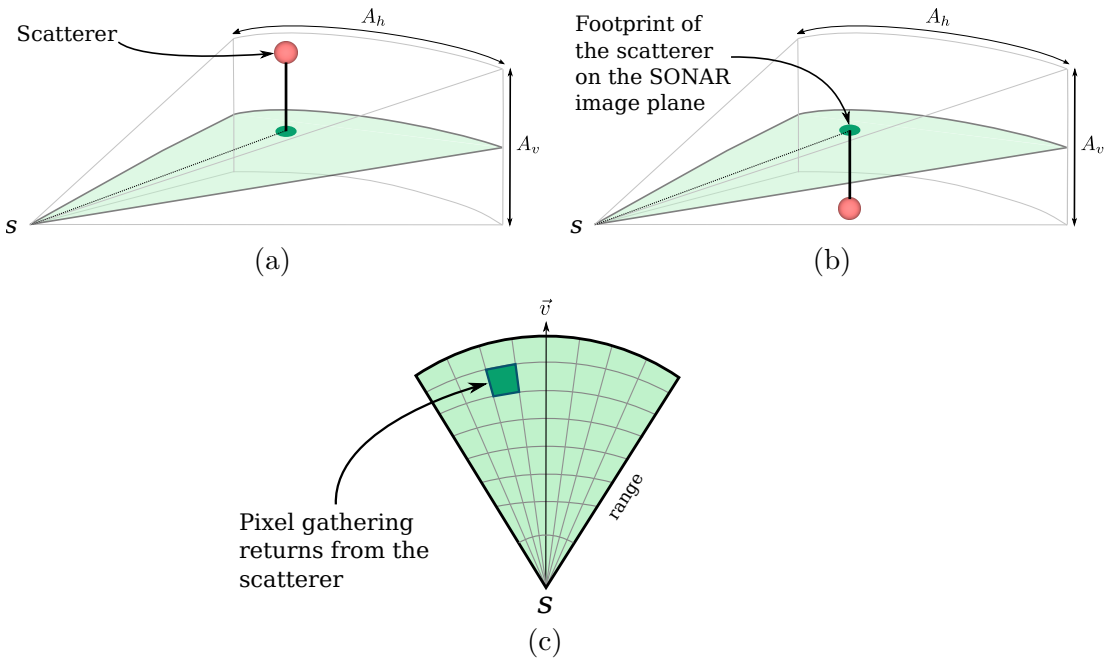



FIGURE 2.15: Illustration of the aperture problem. Two scatterers placed at opposite elevation angles (a and b) lead to the same image (c). In this situation, the returns coming from the scatterers are integrated in the same pixel of the SONAR image with the same measured intensity.

## 2.5 Simulation

In order to simulate realistic underwater scenarios, an underwater simulation environment is needed with the possibility to simulate 2D SONAR data. We present in this section an overview of the underwater simulation possibilities. We then provide a brief review of published work on SONAR simulation and exhibit the limitation of existing solutions. Following this, we present our imaging SONAR simulation framework along with a qualitative comparison between real and simulated data.

### 2.5.1 Choice of the simulation environment

#### Review

In order to provide a realistic 3D underwater environment, sensor simulation algorithms need to be integrated in a simulation environment. A review of commercial and open-source unmanned vehicle simulators is proposed in Craighead et al. [2007]. The authors highlight the recent trend of using frameworks derived from commercial game engines for real-time physics-based simulation and the presence of many open-source solutions. In

the underwater domain, a high-level and non-exhaustive review of AUV simulators has been presented in Matsebe et al. [2008]. In our case, we are interested in an open-source simulation environment in order to allow easy extension and integration of third-party libraries. The simulation environment needs to offer near real-time simulation of multiple sensors and easy configuration of the 3D scenario. After evaluating the different open-source underwater simulation environments available, decision was made to use UWSim (Prats et al. [2012]).

#### UWSim

The UWSim simulation environment offers the simulation of various sensors (Lab [2012]) such as single and stereo video camera, single-beam range reading, structured light projector, object picker, DVL, IMU, GPS and force sensor. The environment supports the simultaneous simulation of multiple vehicles at a time. Physics simulation is provided through the use of OSG library (Osfield et al. [2004]). As visible in figure 2.16, the simulated scene can be defined by a set of CAD models, allowing easy modelling of various real-life situations such as offshore inspections, archaeological surveys, shipwreck inspection, etc. Importantly, UWSim implementation is open-source, making its extension and maintenance by a third party easy. The simulator makes use of ROS framework where new functionalities can be easily added via the implementation of new software modules called nodes.

### 2.5.2 2D SONAR simulation

#### Review

In order to simulate the SONAR imaging process, various approaches have been investigated. Although popular in many simulation applications, finite elements methods are impractical when considering small wavelengths due to a very high computing cost. Frequency domain approaches have been used such as the SIGMAS simulator presented in Groen [2006]. Interestingly, this approach enables easy integration of the beamforming process (match filtering) and of the sensor physical apertures. Following this approach, the authors of Coiras et al. [2009] presented a GPU implementation allowing fast simulation of a side-scan of a few hundreds kHz. The Sonar Simulation Toolset, a ray-based SONAR modelling framework is presented in Goddard [2008]. This framework focuses on fidelity but does not offer real-time data generation. Recently, raytracing-based

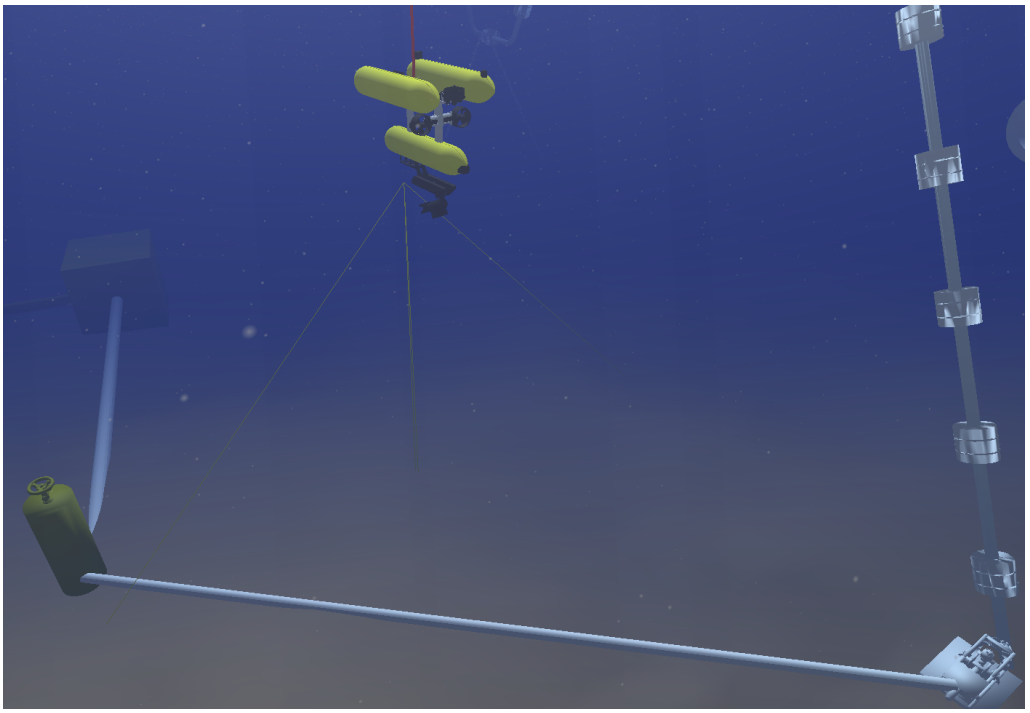


FIGURE 2.16: Simulation of an offshore field with UWSim. A set of CAD models offer realistic 3D representations of oil-field structures.

SONAR simulation methods have been applied to various kind of SONARs [Bell, 1997, Sac et al., 2015] enabling near real-time image generation through discretization of the SONAR beam. A major drawback of ray-tracing approaches applied to SONARs is their discrete and spherical sampling approach which requires high sampling resolutions to avoid subsampling aliasing artefacts. In order to address this issue, improvements of raytracing techniques have been presented in Gueriot et al. [2007] where a so-called tube-tracing technique is used, a tube being defined as a 3D region bounded by an ensemble of rays. Using a tube-tracing technique therefore enables to sample volumes instead of points at a slightly higher computing resource cost. Interestingly, the authors adopter a transducer-wise simulation enabling the simulation of multi-path effects as well as textures due to features smaller than the pixel resolution. Although these raytracing simulators offer realistic intensity shading, no real-time processing is demonstrated. In addition to this, no sensor-specific noise pattern is modelled.

#### 2D Imaging SONAR simulation

In its base version, UWSim only provides very basic multibeam simulation, featuring no vertical aperture modelling. The multibeam readings are then essentially multiple

range readings with no integration along the elevation direction. In order to allow for simulation of imaging SONARs of any aperture, we implemented a SONAR simulation incorporating the imaging model described in 2.4. The implementation is based on a raytracing technique where the SONAR footprint is discretized in an ensemble of rays $\{\wp_{ij} = \wp(\phi_i, \theta_j)\}_{\substack{1 \leq i \leq N \\ 1 \leq j \leq M}}$ at different bearing angles $\phi_i$ (along the horizontal aperture) and elevation angles (along the vertical aperture) $\theta_j$ such that $N = \frac{A_h}{\phi_{res}}$ and $M = \frac{A_v}{\theta_{res}}$ with $A_h$ and $A_v$ the respective horizontal and vertical apertures of the sensor, $\phi_{res}$ and $\theta_{res}$ the angular resolutions considered for sampling the scene. In order to prevent aliasing issues, a jittered sampling (Dippé and Wold [1985]) approach is followed by defining each angle as following a uniform distribution:

$$\phi_i \sim U(-\frac{A_h}{2} + (i-1).\phi_{res}, -\frac{A_h}{2} + i.\phi_{res}) \tag{2.14}$$

$$\theta_j \sim U(-\frac{A_v}{2} + (j-1).\phi_{res}, -\frac{A_v}{2} + j.\phi_{res}) \tag{2.15}$$

where $U(a, b)$ is the uniform distribution on the interval $[a, b]$.

Each ray is then defined by a normalized vector going from the sensor acoustic centre in the direction $(\phi_i, \theta_j)$ w.r.t. the sensor frame. Each ray is then thrown at the scene using OSG library raytracing functions. In case of intersection with an element in the scene, OSG provides the 3D coordinates of the intersection point associated with the local normal which is used to compute the reflected intensity following Lambert law as detailed in equation 2.6. Depending on the value of the elevation angle $\theta_j$, an attenuation factor is applied using a Bessel function as in equation 2.8. The resulting intensity is then accumulated in a $N \times P$ polar image in the corresponding pixel at $(r, \phi)$ following a discretization in the range domain with each pixel being of size $r_{res} \times \phi_{res}$ and $r_{res} = \frac{r_{max} - r_{min}}{P}$. Once all the rays thrown, a predefined scaling of the intensity is applied based on the final quantization of the image (8 or 16 bits) and the maximum level measurable (akin to the physical sensor combined gain in emission and reception).

**Noise simulation**

Once the raytracing image generated, various noise models can be overlaid to provide sensor-specific simulation. In the case of multiple noise sources, the noise distribution often tends to be Gaussian. In this situation, the noise pattern can be measured using a sequence of SONAR images recorded in open water. This sequence can then be used to estimate the first two moments of the noise distribution $(\mu, \sigma)$ at each pixel in the SONAR

(a)

(b)

(c)

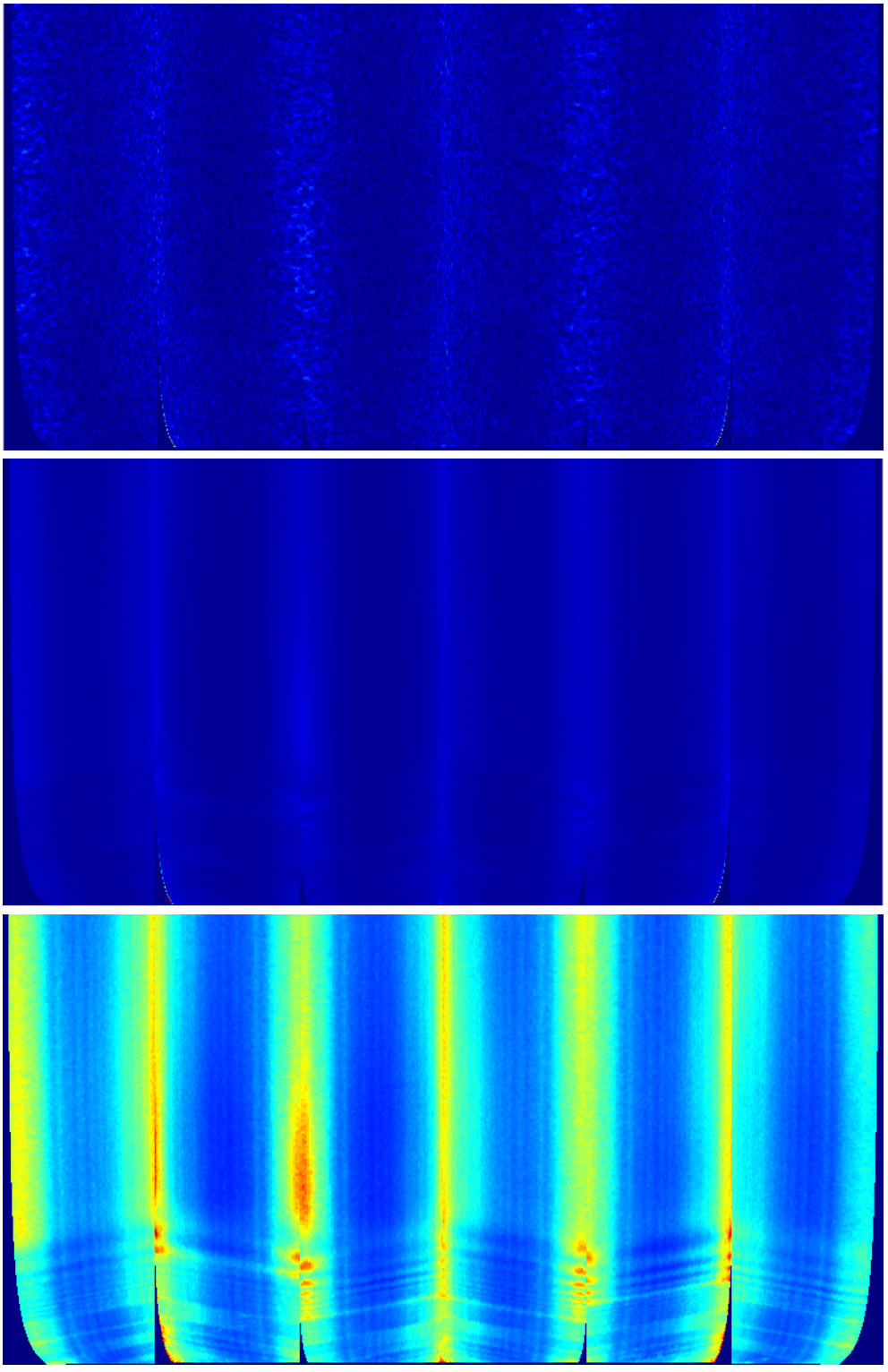


FIGURE 2.17: SONAR noise measurement from a sequence of open-water images. The images are displayed in polar representation. a) Single SONAR image. b) Mean noise intensity values. c) Standard deviation of the noise intensity values.

image. Through this process, two images can be generated, one representing the mean noise intensities $m(r, \phi)$ and the second one representing the standard deviation $\sigma(r, \phi)$ of the noise intensities. As an illustration, figure 2.17 presents a noise measurement using BlueView images acquired in an empty scene in the sea. The images are displayed in polar coordinates (range $\times$ bearing angle). 700 images have been used to obtain a robust statistical estimation. The images are displayed in JetMap intensities where red represents the highest intensity and blue the lowest. The images exhibit significant non-uniformity in the noise spatial distribution. 7 vertical noise stripes are clearly visible with gradual fading on their sides. Based on these statistics, random samples can be drawn for each pixel to emulate the sensor typical noise realisation. The noise data can then be added to the simulated image.

In order to feature a smoother noise model, less specific to the sensor, we chose to use the measurements as a reference for modelling the noise in BlueView images. We modelled the noise as 7 noise stripes with Gaussian intensities of varying mean and standard deviation values at each range and bearing angle, the angular intensity fading being obtained by following a Gaussian distribution around the central bearing angle of each noise stripe.

**Simulation results**

We provide here a comparison of simulated and real data acquired in similar conditions. As illustrated in figure 2.18, the SONAR is mounted in a forward-looking configuration, facing a vertical pipeline. The red line depicts the vehicle trajectory during the inspection of the structure. The green lines delimit the sensor aperture in horizontal and vertical directions. The simulated SONAR features a vertical aperture of 20° and a horizontal aperture (or field of view) of 130°. The distance between the sensor and the structure is of 3 metres, the range resolution is 2cm and the bearing resolution is 0.2° while the elevation angle resolution is 0.01°. The SONAR maximum range is set to 7 metres. Figure 2.19-a shows the output of the raytracing algorithm where intensities due to the beam pattern and Lambert's law are evaluated. Figure 2.19-b depicts the same image with added BlueView-type noise pattern. Figure 2.19-c exhibits a real BlueView P900-130 image acquired in similar conditions, facing a vertical riser featuring advanced marine growth, therefore having a more complex geometry. In spite of this, the images between figures 2.19-a and 2.19-c exhibits similar intensity shading and noise.

The importance of jittered sampling is visible when observing a surface at low grazing

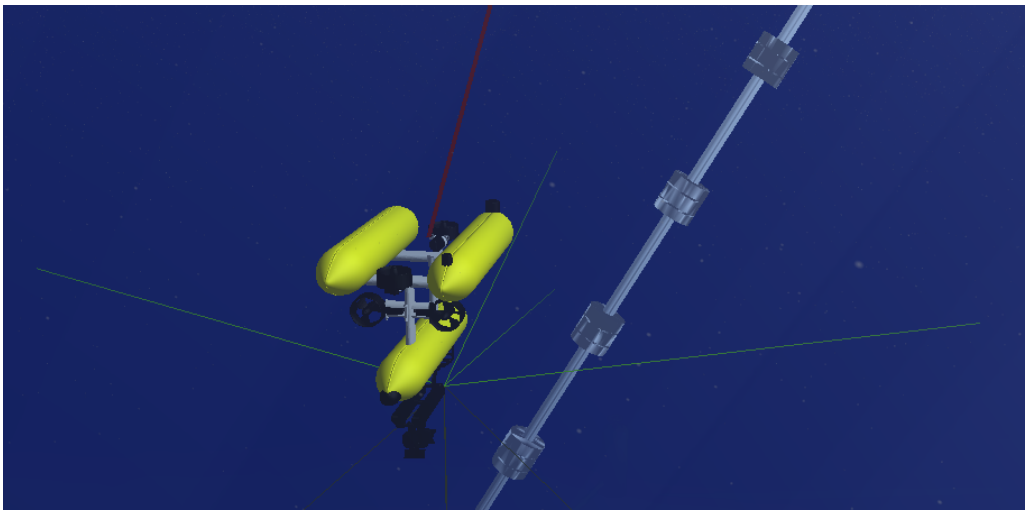


FIGURE 2.18: Simulated pipeline inspection for forward-looking SONAR simulation. As visible with the green lines, the simulated SONAR features a 130° horizontal aperture and 20° vertical aperture.

angles. Figure 2.20 illustrates this on the observation of a flat surface (top side of a cube) with a SONAR. Figure 2.20-a shows the scene to be raytraced by the SONAR simulation algorithm. We present two simulated images, generated at the same spatial resolution and from the same point of view. When sampling on a regular grid, visible aliasing appears on the simulated image as shown in Figure 2.20-b. As visible in figure 2.20-c, this phenomenon disappears when using jittered sampling. On a recent hardware platform (Intel Core i7-4700MQ processor), these images were generated at an average rate of 2Hz, enabling near real-time SONAR simulation. The raytracing process was not multithreaded and the memory usage was limited to a few hundreds MB of RAM, making its integration on small hardware platforms possible.

(a)

(b)

(c)

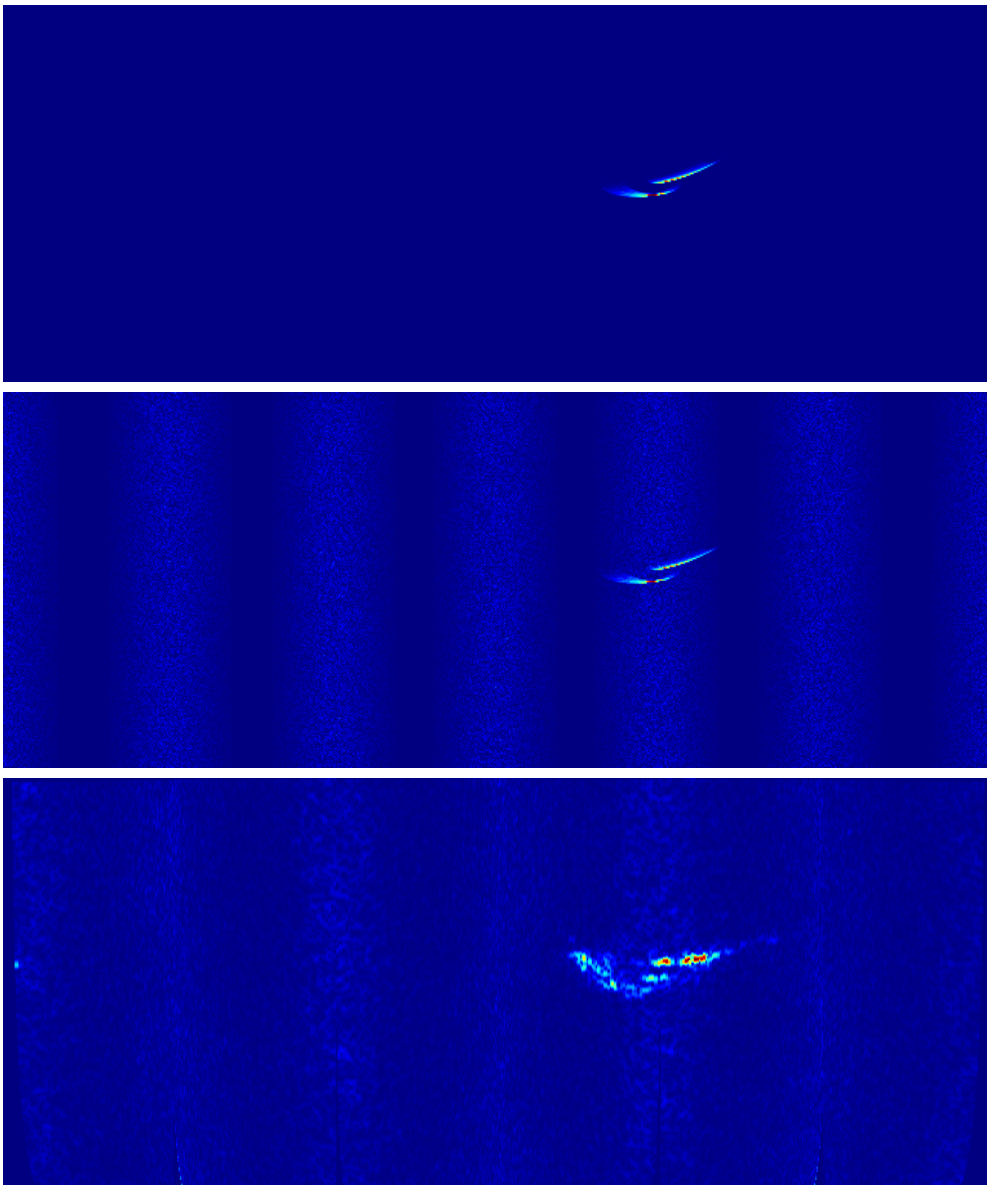


FIGURE 2.19: Comparaison between simulated SONAR data and real data a) Simulated SONAR image with no noise. b) Simulated SONAR image with BlueView-type noise pattern. c) Real BlueView P900-130 image.

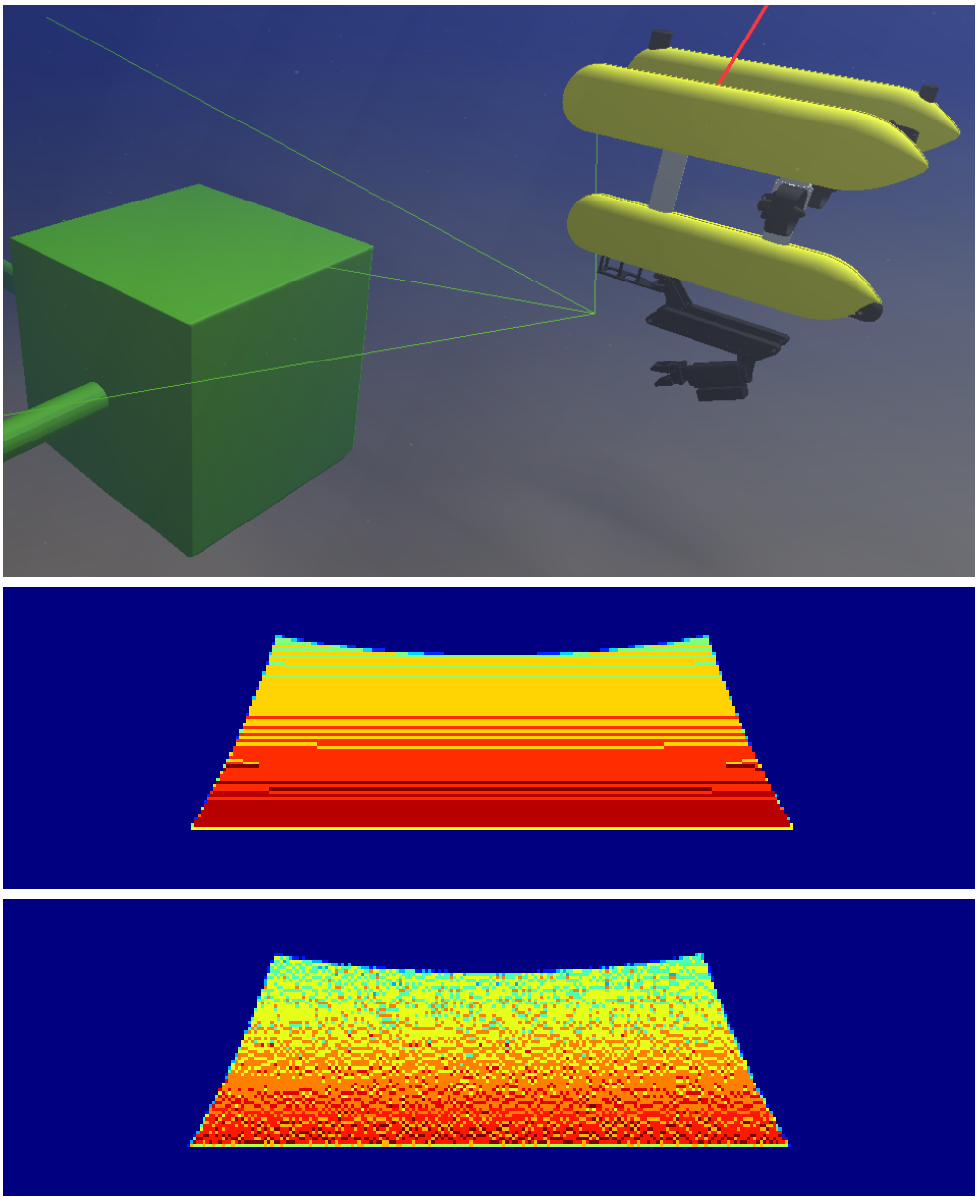



FIGURE 2.20: Illustration of the importance of jittered sampling to prevent aliasing. a) Observation of a surface at low grazing angle. b) Simulated SONAR image with regular sampling, exhibiting visible aliasing. c) Simulated SONAR image with jittered sampling.

## 2.6 Conclusions

In this chapter, we first detailed the principles of SONAR sensing. In order to detail these principles, the various physical phenomena involved in acoustic sensing have been modelled through the description of the emission, the propagation and the reception processes. At each step, a generic mathematical model was presented, deriving the model of a simple acoustic transducer, suitable to describe any type of SONAR.

We presented a review of the different types of SONARs as well as their applications. Importantly, we exhibited the interest of studying the use of 2D imaging SONARS for 3D reconstruction of the scene due to their contained cost compared to 3D SONARS, their large footprints and short to large range sensing capability. Thanks to these characteristics, 2D imaging SONARs are often the sensor chosen to provide sensing in many fundamental applications such as collision avoidance, mapping, target tracking or seabed observation. As a consequence, these sensors are frequently integrated on surface and underwater platforms which makes the acquisition of experimental data easier and the impact of advances in imaging SONAR-based techniques direct.

In light of this, a mathematical model for 2D imaging SONARs was then presented, associating the various elementary models described previously with the specificities of 2D imaging sensors. In particular, the final imaging model exhibits a dependency on the incidence angle to the scatterers surface as well as an integration process along the vertical aperture of the sensor. From these observations result two main consequences. Firstly, the appearance of a scene in SONAR measurements is highly variable and dependent on the position of the sensor. Secondly, the 3D to 2D compression through an integration process makes the estimation of the 3D position of a scatterer along the vertical aperture inherently ill-posed. We illustrated this so-called aperture problem with an example where two different situations lead to a similar SONAR measurement to show the *a-priori* impossibility to obtain 3D information from a single image.

Finally, we introduced our 2D imaging SONAR simulation framework based on UWSim, an open-source underwater simulation environment. Based on the imaging model presented previously in this chapter, this framework provided the capability of imaging any simulated scene with sensor-specific noise model based on a previous modelling of the sensor noise characteristics. We provided details on the implementation based on a

raytracing method as well as a comparison between simulated and real data and showed that unlike existing SONAR simulation solutions, our simulation provides near real-time simulation with realistic noise model.

# Chapter 3

# Literature review on 3D reconstruction from SONAR data

## 3.1 Introduction

Thanks to favourable propagation properties in water, SONARs have been widely employed for reliable range measurement in multiple dimensions. The ability to obtain an accurate 3D representation of the environment is of prime importance in many applications such as mission planning, underwater navigation, environment monitoring, mine countermeasure, archaeology, marine structure maintenance, cave mapping or fishing.

The increasingly large offer of commercially available sensors enabled the development of many scientific and industrial research projects. We provide here a review and analysis of the latest advances in 3D reconstruction from SONAR sensors. In order to highlight the specificities and potential of each type of sensor, we chose to classify our review in five categories representing different types of sensors. We then present our analysis on the pros and cons of each modality and technique used so as to set the context of our work relatively to the current state of the art.

## 3.2 Single-beam echosounder

Single-beam echosounder provide range readings in a single direction, resulting in a small footprint. Their use for 3D reconstruction therefore requires mechanical steering to cover the full space. This cost efficient solution has been employed to study the behaviour of fish populations from a fixed platform in Gauthier et al. [1997] using a 200kHz echosounder featuring an 11° beam aperture. The study presents quantitative results and exhibits the limit of this method such as the difficulty of discriminating targets due to a large beamwidth and target movement during the scanning process which led to multiple recordings of a same target.

While possible on a fixed platform, the use of mechanically steered echo-sounder is made difficult when placed on a moving platform such as an underwater vehicle, requiring accurate relative positioning between the beams and increasing the stochasticity of the backscattered signal. In order to reduce these effects, series of sequential pings can be averaged as in Snellen et al. [2011] where a single-beam 38 kHz Kongsberg echosounder was used to perform sediment classification from a backscatter model inversion. When averaging the results, a trade-off between robust estimation and spatial accuracy is made, therefore limiting the employability of single-beam echo sounders. Recently, Bichucher et al. [2015] demonstrated the ability to simultaneously obtain a bathymetry and correct for slow navigation drift only using a 600 kHz Teledyne RDI Explorer DVL which makes use of four beams.

In general and for 3D sensing purposes, a small beam width provides better spatial accuracy but require higher sampling rate, therefore slower sensor motion. For this reason, single-beam echosounders are rarely used for 3D reconstruction but rather employed for low-resolution data acquisition.

## 3.3 3D from side-scan SONARs

While technically simple, side-scan SONARs only provide time-based backscatter samples of the 3D scene, effectively loosing the elevation angle. In order to perform the conversion from the so-called slant range data to horizontal range data, a flat seabed approximation is commonly used. In general and in the case of a more complex 3D seabed (prominent

3D features such as large rocks or man-made structures), this assumption does not hold true and additional information is needed to recover the 3D information. For this reason, the combination of sidescan data with an additional bathymetry system has been studied in Gueriot [2000] where bathymetric information acquired by a multi-beam echosounder has been associated to side-scan intensities to generate 3D seabed mosaics with additional texture information. Due to the additional bathymetry acquisition, this approach is in practice considerably more time-consuming and expensive. In absence of additional sensors, the recovery of 3D information from side-scan data is in general ill-posed and some prior information is needed.

### 3.3.1 Shadow based estimation methods

In Reed et al. [2004] the shape of objects lying on a flat seabed was estimated from matching the shadow visible behind the objects to simulated shadows generated from a given set of CAD models. Inspired by the early work of Martin and Aggarwal [1983], the use of acoustic shadows was made in Sun et al. [2008] where a Markov random field helped to classify each pixel in one of the three following categories: shadow (no backscattered energy), echo (strong reflection) or background (low-intensity reflection). From this segmentation, a set of occluding contours from different points of views is obtained and used to bound the volume and recover the height information. A reflection map can then be computed by tomography. Results are obtained from simulated data. While interesting, this method requires a circular sampling around the object of interest, which in the case of a side-scan SONAR is very impractical and inefficient in terms of time and surface coverage.

### 3.3.2 Intensity model-based methods

Following the SONAR model, the backscattered acoustic intensity is proportional to the incidence angle of the acoustic wave on the local surface. Based on this observation, shape-from-shading techniques aim at deducing the 3D geometry by inverting the intensity model. Early attempts as in Langer and Hebert [1991] were propagation based: assuming initial values for altitude and slope at the first return, the slope of the following pixels was estimated from the intensity and the elevation values deduced from the slope. Although simple, this technique exhibits a few drawbacks such as error propagation with growing

range, assumed connectivity between successive pixels. The assumed reflectivity model being often non-linear, Durá et al. [2004] presented a linear approach to the shape-from-shading technique, taking advantage of a linear frequency domain model as derived in Bell et al. [1999]. Maximum likelihood estimation approaches (such as energy minimization) aim at matching a model to the observed data by searching for the configuration that minimizes the distance between the model and the data. In Coiras et al. [2007], a side-scan SONAR intensity model is formulated and matched to the observation by the use of an likelihood-maximization optimization procedure. The optimization enables the estimation of the model parameters (seabed reflectivity, side-scan beam pattern and seabed altitude), effectively providing a direct 3D representation of the seabed as shown in figure 3.1. An extension of this technique for arbitrary vehicle motion has been

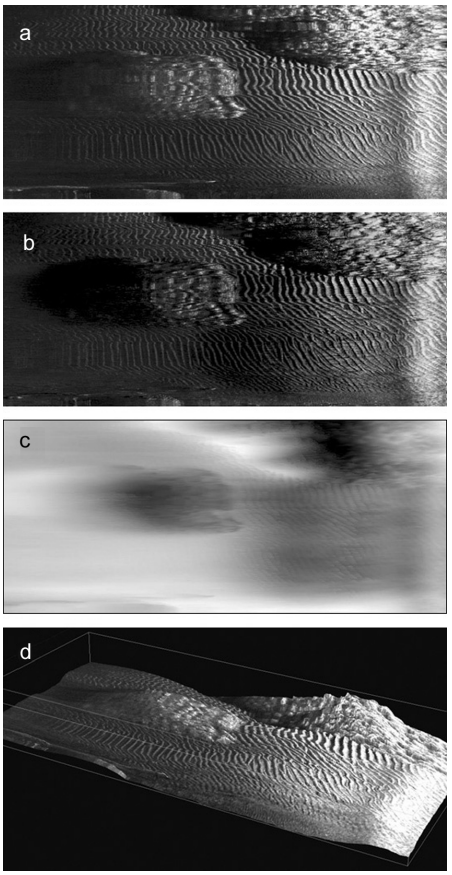



FIGURE 3.1: Illustration of a 3D reconstruction from side-scan SONAR - Coiras et al. [2007]. a) Original side-scan image. b) Model after convergence. c) Elevation map. d) Corresponding perspective view of the textured 3D surface.

presented in Woock [2011].

### 3.3.3 Geometrical model-based method

Depending on the operating scenario, side-scan measurements can be affected by well-known multi-path effects, effectively integrating reflections coming from multiple directions such as a fish, the seabed or the sea surface. As shown in Saucan et al. [2015], the ability to predict the DOA (Direction Of Arrival) enables 3D reconstructions at a much higher level of detail. The authors present a so-called echo-tracking method based on a set of geometrical models expected in the scene. Each measurement is tested against all models and a tracking approach helps to regularize the observations. As can be seen in figure 3.2, experimental results from real data exhibit accurate 3D reconstruction.

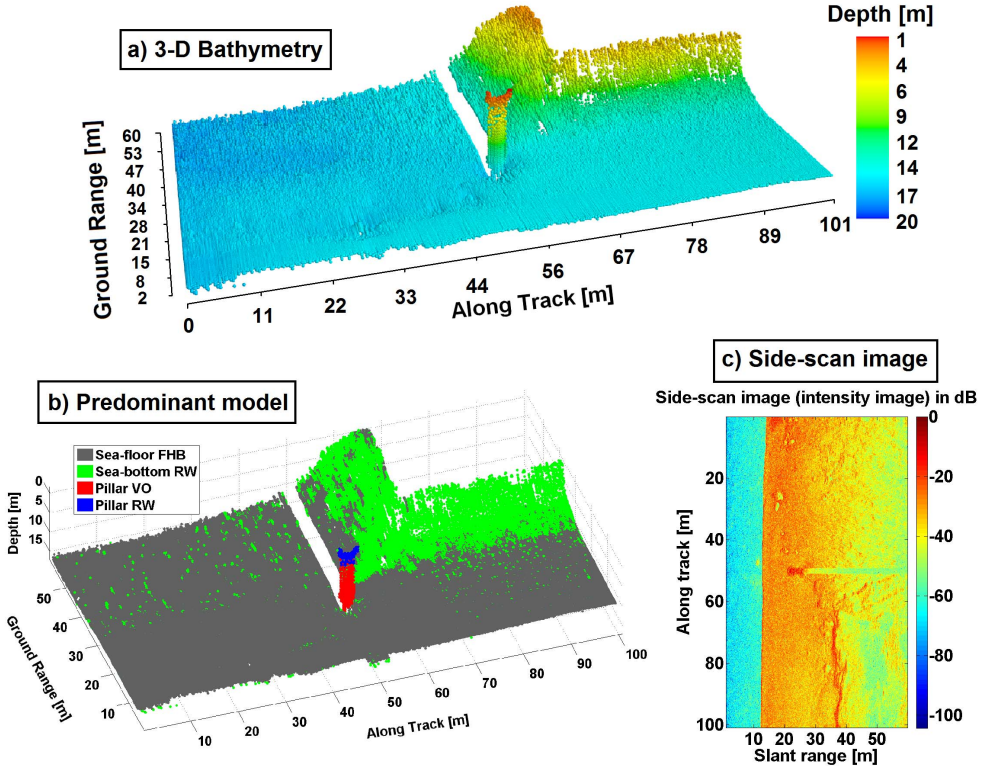



FIGURE 3.2: Illustration of a 3D reconstruction from side-scan SONAR with model prediction - Saucan et al. [2015]. a) Reconstructed bathymetry. b) Selected model. c) Original side-scan image.

### 3.3.4 3D side-scan

Inspired by the principle of SAS (Synthetical Aperture SONAR) where the angle of reflected waves is estimated by interferometry, the authors of Griffiths et al. [1997]

presented a 3D side-scan system. The prototype sensor has been tested in a controlled environment, yielding low resolution reconstruction. The main reason for the limited accuracy of this configuration is well-know in the SAS community, as these systems require very accurate knowledge of the sensor position (to a fraction of the wavelength) equivalent at best to mm-level navigation accuracy which is currently impossible to obtain on underwater platforms.

For this reason, this technology has only been punctually applied, both commercially as in Hartley et al. [1993] and for research purposes (Sæbø et al. [2013]) as illustrated in figure 3.3.

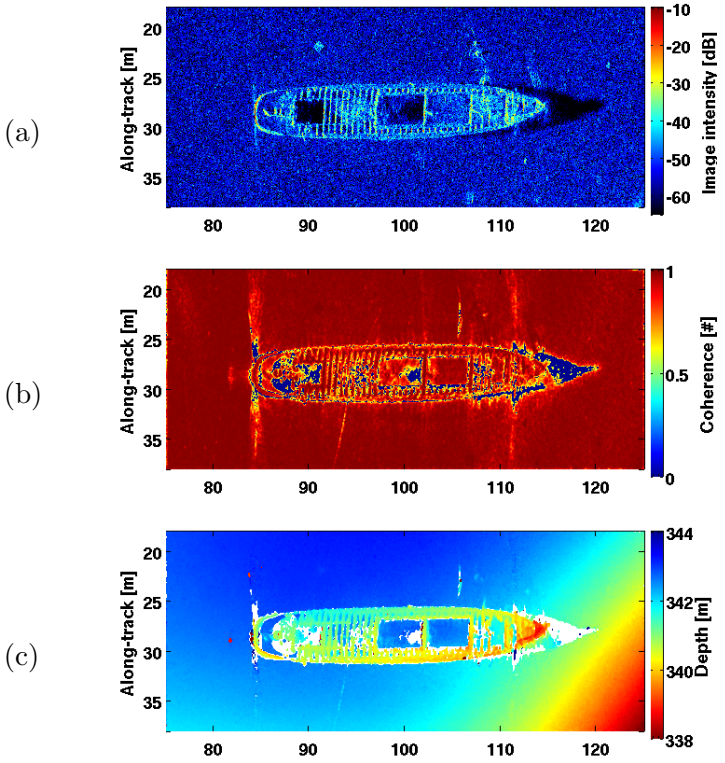
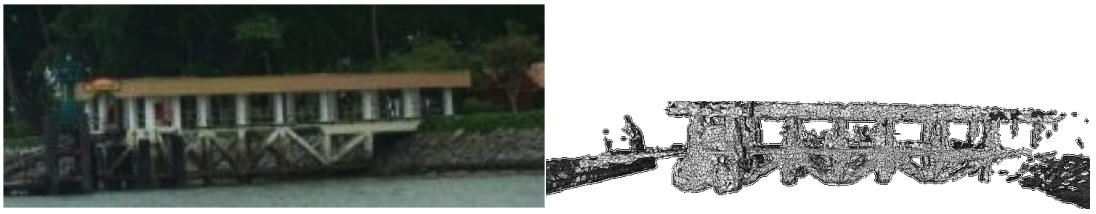



FIGURE 3.3: Illustration of a interferometric SAS system - Sæbø et al. [2013]. a) SAS image. b) Interferometric coherence image. c) Estimated bathymetry using a complex cross correlation technique.

Although widely used for their long-range imaging capabilities, the operation of side-scan SONARs for 3D reconstruction purposes remains marginal and is based on the assumption that each range reading corresponds to a single altitude on the seabed, therefore restricting it to the observation of simple and smooth surfaces. However, in these situations, good results have been obtained by taking advantage of an accurate intensity model and adequate inversion methods.

## 3.4 3D from pencil-beam imaging SONARs

Nowadays commonly used for the study of ecosystems and fishing (Gerlotto et al. [2000]), pencil-beam multibeam SONARs offer accurate range and bearing readings thanks to their unidimensional array of transducers. This results in accurate 2D sensing with low uncertainty along the remaining dimension due to very small apertures (typically 0.5° to 1.5°). Direct 3D sensing can then be obtained by combining the 2D pencil-beam SONAR images with a motion along the missing direction. The reconstruction procedure in this case is a simple projection of the range profile represented in the acquired image. This profile which is essentially an ensemble of range readings (one per bearing) is typically obtained by selecting the first return (value higher than a given threshold) as in Newman and Durrant-Whyte [1998] or the maximum intensity along the beam. As a result, this configuration is widely used to perform underwater 3D sensing at a range of up to 100 metres as presented in Gerlotto et al. [1999] where the authors applied this technique to the observation of fish schools in 3D. Since their first non-military application (Farr [1980]), pencil-beam SONARs have been frequently used to acquire bathymetry data both from the surface when mounted on a boat (Vaneck et al. [1996]) or an AUV (Grasmueck et al. [2006]). In this context, the accuracy of the map is usually limited by the accuracy of the navigation data but multiple research efforts showed that accurate maps could be obtained when combining the acquisition of multi-beam data with a SLAM approach as in [Barkby et al., 2011, Palomer et al., 2016, Roman and Singh, 2005]. Cave reconstruction was explored by Mallios et al. [2015] using a mechanically scanned imaging sonar with a 1° beam width. A high-frequency multi-beam SONAR, a BlueView MB-2250, has been used in Papadopoulos et al. [2011] to map a marine structure simultaneously underwater and above water using a LIDAR sensor (see figure 3.4). Interestingly, no navigation sensor such as DVL or GPS was used to position the acquired data. The authors performed the mapping using a scan registration technique from LIDAR data and presented a 10cm resolution 3D reconstruction. To the exception of this study, multibeam-based 3D reconstructions require navigation data to enable multiviews data association. Conversely when operating at low frequency, acoustic waves penetrate further in the seabed. Low-frequency multi-beam SONARs like the Kongsberg Simrad EM1002S can therefore provide both bathymetry and backscatter information as detailed in Brown and Blondel [2009]. In [am Ende, 2001, Stone et al., 2000], cave mapping was achieved using the DWM (Digital Wall Mapper), a diver-assisted



(a) (b)

FIGURE 3.4: Illustration of a 3D reconstruction from pencil-beam SONAR and LIDAR data - Papadopoulos et al. [2011]. a) Reconstructed structure. b) Reconstructed surface.

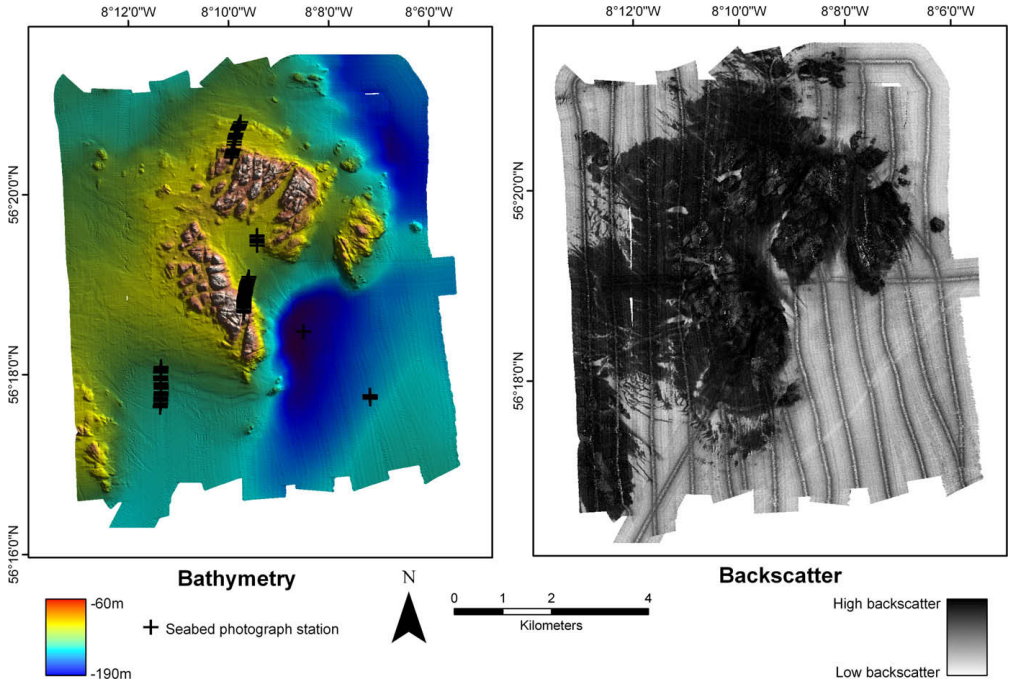
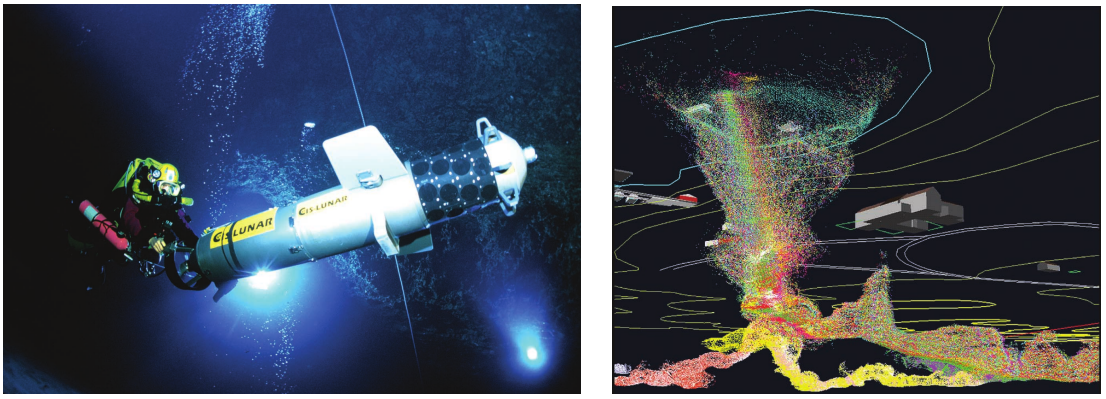



FIGURE 3.5: Illustration of bathymetry and backscatter data obtained with a pencil-beam SONAR - Brown and Blondel [2009]. The bathymetry (visible on the left side) and the backscatter data (visible on the right side) are both acquired from a single pencil-beam SONAR, the Simrad EM1002S.

torpedo-shaped instrument featuring acoustic sensing capability as well as a propeller. The mapping was done using thirty-two 2° aperture transducers, helically arrayed around the nose of the instrument. After manual registration of the acquired point clouds, results exhibit good large-scale mapping ability with a 21km long mapped dataset as can be seen in 3.6.

Similarly, a set of 54 pencil-beam transducers has been used in Fairfield et al. [2007] to map underwater tunnels. While this modality provides direct 360° mapping capability, it requires specific integration on the vehicle.




(a) (b)

FIGURE 3.6: Illustration of a 3D cave mapping from an array of narrow beam transducers - am Ende [2001]. a) The Digital Wall Mapper instrument, featuring 32 helically arrayed narrow beam transducers. b) Illustration of a 3D reconstruction of the entrance of a cave.

Due to their small footprint, pencil-beam imaging SONARs have been widely used for bathymetric measurements and short range marine structure reconstruction. A trade-off is then achieved in the design of the sensor to provide either high range accuracy with low maximum range when operating at high frequency or on the contrary lower range accuracy but longer maximum range when operating at low frequency.

## 3.5 3D from wide-beam imaging SONARs

Unlike pencil-beam SONARs, wide-beam imaging SONARs provide typical apertures of 7° to 20°. Wide-beam imaging SONARs therefore ensonify large volumes of water at a time, providing information on large amounts of 3D points in the observed scene.

Due to their larger aperture, wide-beam SONARs produce integral intensity values that result from a summation along the vertical aperture, effectively producing a blurring effect. Each measurement therefore potentially corresponds to reflections coming from multiple points in the 3D scene. The 3D reconstruction process is therefore an ill-posed problem and requires additional information to be solved.

### 3.5.1 Non linear methods

In a similar way to reconstruction from side-scan data, various approaches have been investigated such as the use of acoustic shadows in Zerr and Stage [1996] where the authors reconstructed a small object placed on the ground from a set of 2D images. These images were acquired from a set of positions located on a circle around the object, separated by 5° steps. For each image, the pixels were classified into 3 categories (echo, shadow or background) using a Markov random field and intensity thresholds. The shadows and sensor positions were then used to bound the object volume in each point of view. Good results were obtained in a controlled environment featuring a flat ground where a 360° rotation around the object was possible.

Similarly, Aykin and Negahdaripour [2013, 2016] evaluated the reconstruction of objects laid on a seafloor from multiple views around the objects taken at both multiple yaw angles and roll angles. In this situation, the shape of the observed is obtained by successive volume bounding from a background, return, shadow segmentation. The roll angle in the sensor enables to take advantage of the high bearing resolution of the sensor to compensate for the uncertainty along the vertical axis. Not only this method assumed the presence of a background surface to use the shadow information but it also required roll motions and a large rotation around the object which are both impractical when inspecting real marine environments with an AUV.

### 3.5.2 Imaging model inversion

In Aykin and Negahdaripour [2013], the SONAR imaging model is inverted to recover the missing elevation of range readings from the measured acoustic intensity. Controlled environment experiments are presented, exhibiting reconstructions of small objects with limited accuracy and visible distortion in the vertical direction. Due to the impossibility of computing the elevation behind acoustic shadows, this work is also based on the assumption that the reconstructed objects feature smooth surfaces, varying monotonically in terms of distance from the sonar which does not hold true in general.

### 3.5.3 Opti-acoustic methods

As reviewed in Ferreira et al. [2016], recent improvements in SONAR sensors accuracy paved the way to easier association with optical sensors. While SONARs provide accurate range measurement and uncertain elevation information, optical cameras offer by design high angular resolution in elevation and bearing but no depth information. Combining these two sensors therefore enables instant 3D sensing. When using two different sensors, arises the need of joint calibration to ensure geometrical consistency when performing 3D reconstructing. In this situation, standard calibration such as the Direct Linear Transform (Hartley and Zisserman [2003]) appear impractical due to the difference in spatial resolution and acoustic noise. As detailed in Hurtós et al. [2010], alternative calibration methods can be employed to achieve consistent mapping. In Babaee and Negahdaripour [2015], a so-called opti-acoustic imaging system has been used to provide observations both in the optical and in the acoustic domain. Through the use of 2D occluding contour correspondences, a 3D occluding rim is defined and opti-acoustic samples are registered by bundle adjustment. From this, a simple surface interpolation was employed to reconstruct the surface of small coral rocks in a controlled environment. As can be observed in figure 3.7, the reconstruction results exhibit decent accuracy but require large rotations around the objects and assume contours visibility in both sensors which implies in particular, having the SONAR pointed in a grazing configuration and low water turbidity to allow for sharp optical imaging.

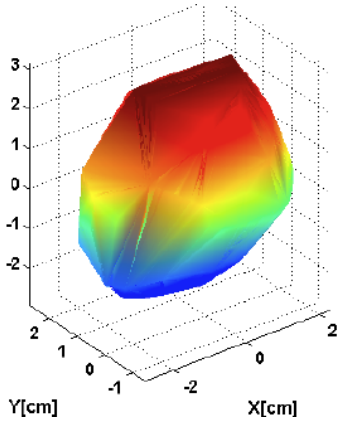



FIGURE 3.7: Illustration of reconstructed stone from combined multiple wide-beam SONAR and optical camera observations during a 360° rotation around the stone - Babaee and Negahdaripour [2015].

In Ozog et al. [2015], prior information on the object to be reconstructed was given by either a CAD model or a video-based 3D reconstruction. A Didson imaging SONAR was used for its large footprint when projected on a ship hull section. The prior map was projected in the SONAR image plane to associate the 3D position to the acoustic measurement in the SONAR image, based on the known sensor position. More precisely, for each mesh element of the prior map, the four best views were selected and the acoustic intensities were blended based on their distance to the mesh sampled points. As illustrated in figure 3.8, this results in a large 3D map, textured with acoustic intensities. While this technique exhibits interesting results at large scale, no small scale error analysis is made, due to the lack of ground truth. Furthermore, this approach is entirely based on the use of an accurate prior map which is in general hard to obtain.

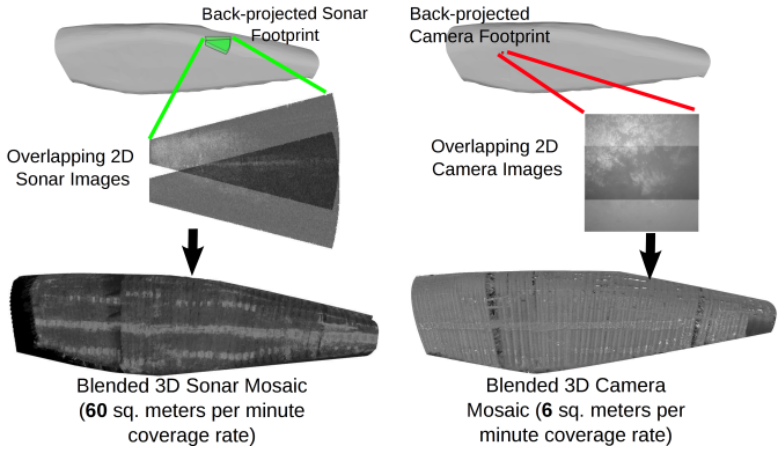



FIGURE 3.8: Illustration of ship hull mosaic obtained with a wide-beam imaging SONAR - Ozog et al. [2015]. The back-projected SONAR footprint appears much larger than the camera footprint, enabling faster mosaicing.

### 3.5.4 Feature-based methods

Recent work (Huang and Kaess [2015]) investigated the application of structure from motion techniques to Didson SONAR data using manually selected and associated feature points. While interesting, this approach relies on the availability of good acoustic features and accurate associations. In SONAR imagery, the variability of the appearance of 3D points based on the position of the sensor is well known. In addition to this, measurement noise and frequent acoustic effects such as multi-path ringing make the observed intensity prone to high variations from one point of view to another. As a consequence, a fully

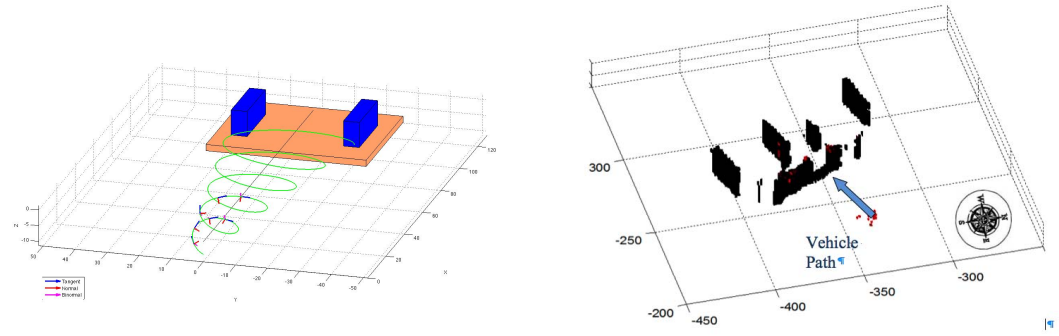
automated processing with acoustic feature computation and an automated corresponding association is likely to be unreliable, making acoustic structure from motion impractical.

In a similar way and inspired by computer vision techniques, Brahim et al. [2011] presented a reconstruction technique based on stereo-matching between two SONAR images. Building up on their previous work on feature point extraction (Brahim et al. [2010]) which presents a feature extraction method based on multiple frames to reduce speckle noise, the author detailed an assessment of the Didson SONAR projective model and a 3D reconstruction method based on a evolutionary optimization algorithm. The results exhibit very sparse 3D reconstruction, only allowing for basic structural representation featuring the corners of the object. Although sparse, feature-based methods enable navigation-free registration through feature matching.

### 3.5.5 Acoustic stereo imaging

In Assalih et al. [2013], ASI (Acoustic Stereo Imaging) has been used to obtain real-time 3D reconstruction from two imaging SONARs. Knowing the transformation between the two sensors, the 3D position can be estimated by triangulation. The results from an experiment made in a controlled environment exhibit potential good accuracy but high variability in the results. SONARs being expensive sensors, embedding two sensors on a vehicle is impractical but assuming the scene is static and good position information, one moving sensor could achieve the same results. The remaining problem lies in data association across multiple points of view when dealing with uncertain positions which has not been addressed in this study.

Similarly, the use of two SONARs has been investigated in Horner et al. [2009] where two different sensors were mounted in orthogonal directions, providing a direct way to obtain the 3D coordinates of the points lying at the intersection of the two footprints. Through the use of a Bayesian filtering, an occupancy grid is iteratively built, providing an online obstacle avoidance method. Experimental results are demonstrated on a navigation sequence under a bridge, demonstrating a path-planning-free navigation in presence of obstacles. As illustrated in 3.9, only low-resolution mapping is achieved and comes at the cost of a helical trajectory.




(a) (b)

FIGURE 3.9: Illustration of dual wide-beam SONAR 3D occupancy estimation - Horner et al. [2009]. a) The helical trajectory (green) enables to ensonify surrounding hazards while maintaining the desired forward heading. b) Estimated occupancy grid of the scene after reconstruction.

### 3.5.6 Acoustic concentrator lens

A Didson sonar has been used in Mizuno and Asada [2014] combined with a 3° concentrator lens in order to reduce the aperture, enabling direct 3D reconstruction. Motion correction using additional sensors and 3D mosaicing was then performed to create a 3D map of a lake. As presented in figure 3.10, a segmentation was made using a difference of Gaussian methods, enabling the classification of voxels in three categories: seabed and two types of aquatic plants present in the lake (*Chara globularis* and *Elodea muttallii*).

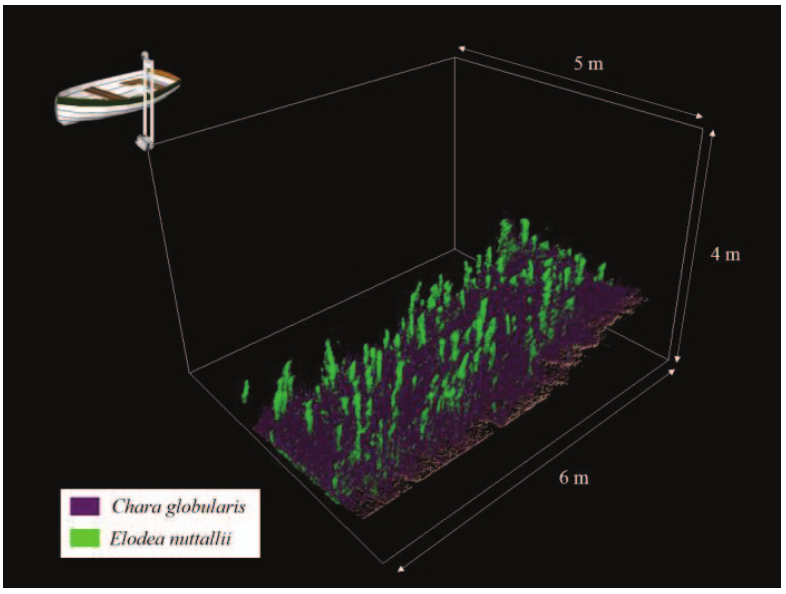



FIGURE 3.10: Illustration of reconstruction of aquatic plants using a 3° concentrator lens mounted on a boat - Mizuno and Asada [2014]. The reconstruction exhibits two different types of plants, *Chara globularis* and *Elodea muttallii*.

## 3.6 3D SONARs

Due to the uncertainty along the elevation axis when using unidimensional arrays, the interest in 2D arrays of hydrophones has grown considerably in the last few years.

### 3.6.1 Bio-inspired 3D SONARs

Inspired by the human eye, Rosenblum et al. [1991] presented a 3D SONAR composed of an acoustic lens focusing the incoming waves on small transducers laid on a half-sphere (see figure 3.11), in a similar way to the human retina. Direct reconstruction can then be obtained and a simple threshold enables the separation of noise and acoustic returns. The authors present results of reconstructed spheres at a voxel resolution of 10cm, mainly limited by the beamwidth.

### 3.6.2 Mechanically scanned SONAR

An inexpensive approach to directly observe the 3D space is to rotate a single-beam along two dimensions. An example of this approach is used in Auran and Malvig

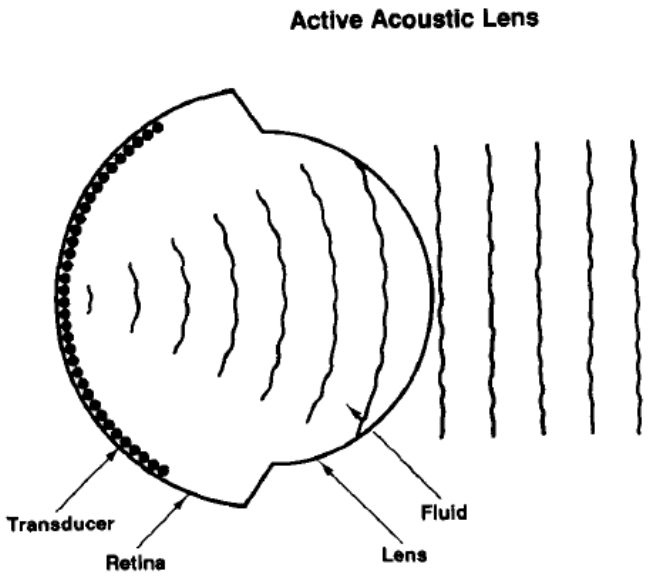



FIGURE 3.11: Illustration of a bio-inspired 3D SONAR - Rosenblum et al. [1991]. An array of transducer receives acoustic waves focused by the acoustic lens.

[1996] where a WesMarSS150 sensor is used, featuring a beam-with of 6.5°. The sonar head could be tilted at a wide range of angular configurations ($\phi \in [86°,180°]$ and $\theta \in [0°,360°]$). The authors present a method to compute connectivity between 3D sonar returns, low-resolution representation is given. One major limitation is the considerably large sampling time (100 seconds per full sweep) requiring to stabilise the platform (or accurately estimate its motion) during the acquisition.

Similar work has been carried on in Roman and Singh [2004] where a pencil beam 2D scanning sonar producing range images was used to measure 3D bathymetry patches. As illustrated in figure 3.12-a, these patches were then registered using a scan registration technique to achieve micro-bathymetric mapping (see figure 3.12-b). Although promising, a significant amount of time is needed to acquire each patch, restricting this technique to applications where the sensor can be kept steady or well positioned. For these reasons, mechanically-scanned sensors are rarely used on moving platforms.

### 3.6.3 Pencil-beam 2D array

In Jaffe et al. [1995], a 3D SONAR made of 64 pencil-beam (2°) transducers is presented. The prototype called FishTV featured a 2° angular resolution and a 16° field of view,

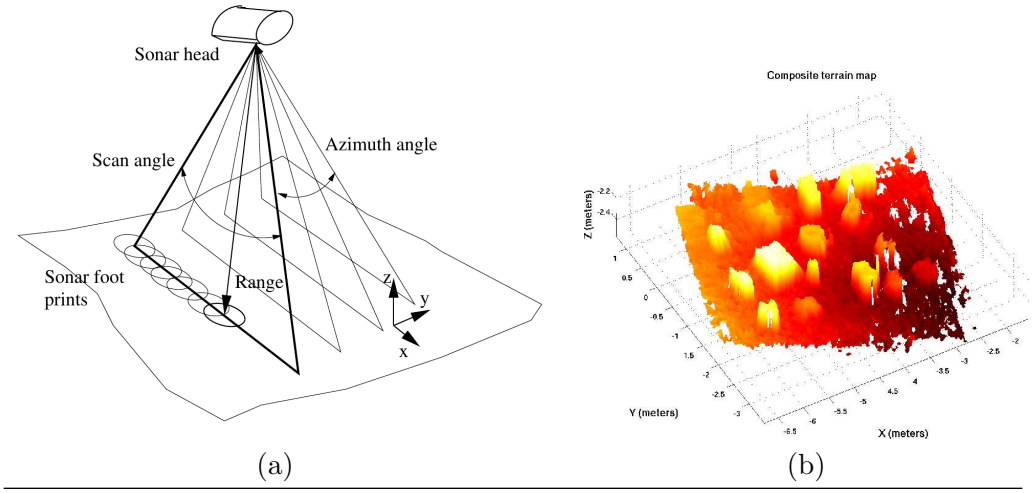


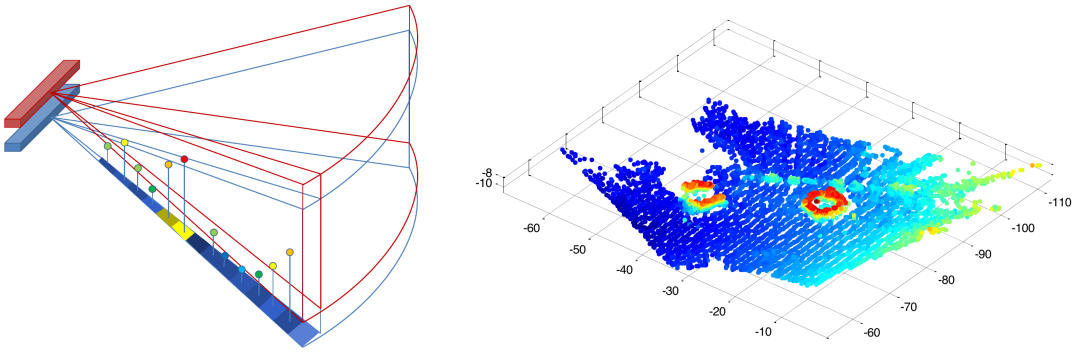

FIGURE 3.12: Illustration of mechanically-scanned bathymetry system - Roman and Singh [2004]. a) A 2D mechanically-swept pencil-beam sensor enables the recovery a 3D patch. b) Illustration of reconstructed terrain bathymetry assembled from 6 patches.

allowing the detection and tracking of zooplankton. Due to the absence of beamforming processing, the resolution is limited to the transducers aperture and restricts its usage to tracking species of a few centimetres size.

### 3.6.4 Interferometric 3D SONARs

In opposition to mechanically scanned SONARs, the steering process can be obtained electronically from a bidimensional array of transducers as presented in Zimmerman [2004]. This method is often refereed to as interferometry. In Yufit and Maillard [2013], a prototype of a so-called FLBS (Forward Looking Bathymetric SONAR) made of three SONAR arrays is presented, allowing the estimation of elevation information by phase difference (see figure 3.13-a). As illustrated in figure 3.13-b, the device allowed for long-range reconstruction (60 metres away from the target) with a resolution of 10cm.

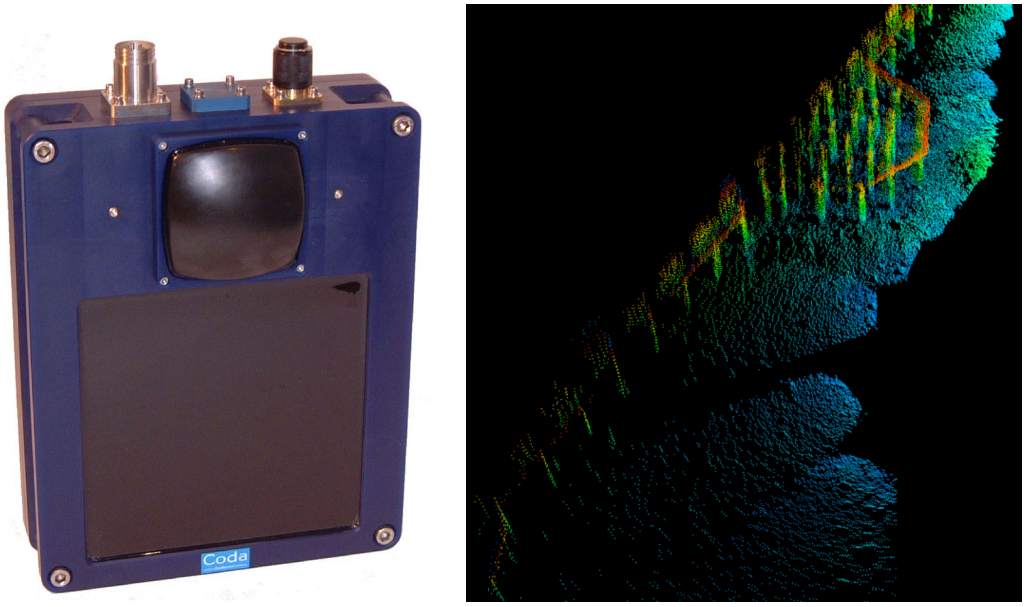
Similar solutions are now commercially available such as the Tritech Eclipse. This sensor was used in Bülow and Birk [2011] where a frequency-based method was presented to register multiple 3D scans. The reconstructed data shows low-resolution 3D information but coarse geometrical consistency. Since its first presentation in 1996 (Hansen and Andersen [1996]) as a 3D acoustic camera, the Echoscope sensor has seen its performance improving. Recently, the Coda Echoscope technology presented in Davis and Lugsdin [2005] showed very promising results (see fig3.14), exhibiting an improved 3D resolution




(a) (b)

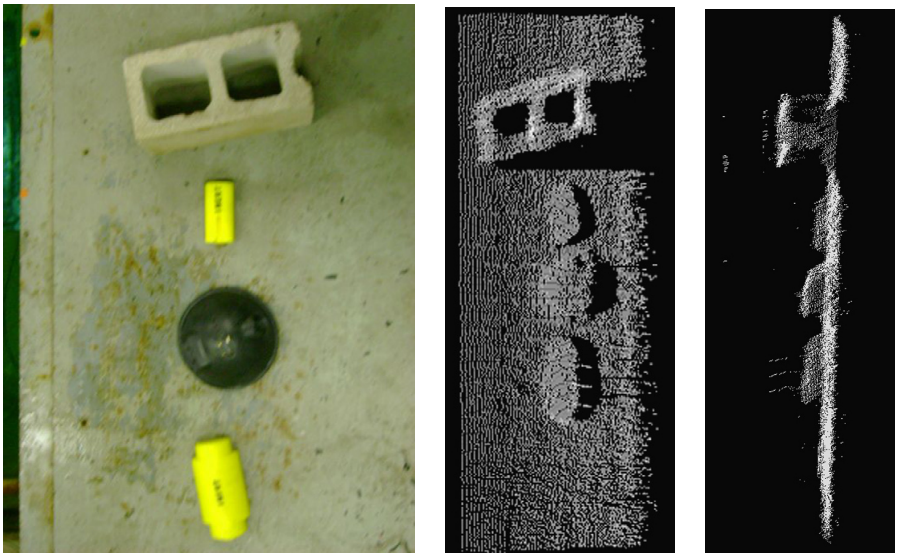
FIGURE 3.13: Illustration of FLBS system - Yufit and Maillard [2013]. a) Elevation estimation of the targets (spheres) is achieved by phase difference measurement with the two receiving arrays of transducers (respectively represented in red and blue). b) Illustration of FLBS long-range reconstruction. The reconstruction exhibits two concrete cylindrical blocs laid on the seabed.

of up to 1cm. Operating at 375kHz, long-range (up to 200 metres) 3D reconstruction can also be achieved. Thanks to its instant 3D sensing capability, this sensor has been used in recent research work such as fast 3D mosaicing of a shipwreck (Hansen et al. [2005]), real-time positioning (Woodward et al. [2010]) or the study of ship wakes (Soloviev et al. [2012]). In order to address the noise in the Echoscope range images, a method based on Markov Random Fields has been applied in Murino et al. [1998] to obtain a measure of confidence on the range readings. An iterative optimization enabled the recovery of centimetre-level accuracy range readings, greatly improving the 3D representation of small objects. In spite of its attractive performance, the Echoscope remains an expensive and relatively large sensor which makes its integration on AUVs or cost-efficient platforms difficult. In order to address the size issue, recent research work have adopted higher frequency designs such as in Josserand and Wolley [2011] where a 8 x 8 cm 2D SONAR array is presented with a spatial resolution of 2.5 cm in range, 1° in azimuth and 1° in elevation. The authors took advantage of a FSPA (Frequency Steered Phased Array) technology enabling beam steering in different directions depending on the applied frequency. Tests in water tank using a broadband 2.25 MHz transducer are presented, showing fairly good accuracy 3D images as can be seen in figure 3.15.




(a) (b)

FIGURE 3.14: Illustration of the Coda Echoscope system - Davis and Lugsdin [2005]. a) The Coda Echoscope sensor is composed of a 2D array of transducers. b) Illustration of a single-pass 3D recontruction obtained with the Echoscope while inspecting a harbour. The pointcloud exhibits multiple pillars and a detailed rocky seabed.



(a) (b) (c)

FIGURE 3.15: Illustration of a high-frequency FSPA SONAR image - Josserand and Wolley [2011]. a) Objects placed in a water tank for reconstruction. b) 3D Image obtained by frequency beam-steering. c) Same 3D image from a side point of view.

## 3.7 Summary

We present in table 3.1 a summary of the different 3D reconstruction methods presented in this chapter and exhibit the advantages and inconvenients of each approach. In the context of observations gathered by and AUV and in spite of their high resolutions, the restrictions on bulkiness prohibits the use of 3D SONARs. The observation of moving elements and the necessity to react to changes in the environment (dynamic path planning) make mechanically steered sensors unpractical. Shadow-based techniques require a background and knowledge on its shape which restricts their use to flat seabed inspections but do not allow for instance the inspection of vertical pipelines. 3D SAS approaches provide high-resolution reconstructions but require mm-level navigation accuracy which when operating at high-frequency is impossible to achieve on AUVs with current navigation systems. The employment of multiple sensors feature the inconvenients of each sensing modality and require a joint calibration. Feature-based attempts exhibited low-resolution reconstructions. As opposed to these methods, model-to-data matching methods exhibit robustness to outliers and noise and have the potential to offer accurate reconstruction provided a good sensing model is available. To the exception of 3D SONARs and feature-based methods, all methods require navigation data (DVL, INS or USBL-based) to estimate the displacement of the sensor between successive views.

| | Pros | Cons |
|---|---|---|
| Direct sensing through mechanical steering | - low cost | - low framerate<br>- limited to slow changing conditions (slow motion of the vehicle and elements in the scene) |
| Shadow-based and shape-from-shading | - can be obtained using respectively low-cost (sidescan) or high-resolution (multibeam) sensors. | - requires the presence of a shadow on a known background (typically assumed to be a flat seabed)<br>- requires accurate imaging model<br>- requires accurate navigation |
| model-to-data matching | - can be obtained using respectively low-cost (sidescan) or high-resolution (multibeam) sensors.<br>- robust to outliers and noise in data (image or navigation) | - requires accurate modelling<br>- optimisation approaches are computionally expensive |
| 3D SAS | - high resolution (cm level) | - requires perfect position estimation (mm-level) |
| multi-sensor (multiple sonars or opti-acoustic) | - can leverage strenghts of multiple modalities such as range resolution for SONARs and elevation/azimuth angles resolution for video cameras | - increased cost due to multiple sensors<br>- requires joint calibration<br>- often limited by disadvantages of both sensors<br>- might require multiple inspections |
| feature-based | - can be performed without navigation input (feature matching) | - low resolution (sparse features)<br>- sensitive to appearance variability of SONAR data |
| direct 2.5D using an 2D transducer array (3D SONAR) | - high resolution (cm level)<br>- does not necessarily require navigation input (registration by 3D matching) | - expensive sensor<br>- bulky |

Table 3.1: Comparison of state-of-the-art 3D reconstruction methods

## 3.8 Conclusions

The problem of 3D reconstruction from SONAR sensors has been studied regularly for now more than three decades, involving different sensor configurations. Recent advances in 2D array processing techniques have paved the way to the development of 3D SONARs, providing direct and real-time 3D information at ranges of up to 200m. Mechanically-steered SONARs offer inexpensive solutions but require accurate position knowledge during the scanning process, mainly restricting their use to steady platforms or low refreshing rate applications. In comparison to these sensors, multiple interferometric 3D SONARs have been successfully developed and commercially used in the last few years. These sensors now offer centimetre-level accuracy but remain rarely embedded on AUV platforms due to their large size and high cost.

Inexpensive and technically simple single-beam sensors such as single-beam echosounders and side-scan SONARs are often embedded on surface and underwater platforms. In spite of their popularity, they remain rarely used for 3D sensing, either due to their small footprint (single-beam echosounder) or to the uncertainty on the angle of arrival in the case of side-scan SONARs.

Since spatial accuracy generally comes at the cost of a small footprint, small-aperture multibeam (or pencil-beam) imaging SONARs offer an interesting trade-off by combining an array of high-frequency transducers with beamforming techniques to generate 2D images with low uncertainty in every 3D direction, at ranges of up to a few tens of metres. These sensors offering direct 3D sensing capabilities, they have been widely used for bathymetric applications and marine structure inspections. The accuracy of the 3D reconstruction is in this case limited by navigation accuracy, restricting its use to local mapping. In the case of large mapping applications, pencil-beam sensors remain of interest but require the use of registration techniques applied between multiples views, increasing the operation time. Furthermore, the small footprint of pencil-beam sensors leads to overall lower scene coverage rate, leaving gaps between scans and unobserved surfaces when observed at grazing angles.

Conversely, wide-aperture imaging SONARs provide larger footprints and are for this reason widely used for seabed imaging and monitoring applications. The ability of scouring large volumes of water at a time comes at the cost of an increased uncertainty

in the elevation direction. In order to solve for this uncertainty, additional equipment such as optical camera, acoustic lenses or a second SONAR has been used, increasing the cost, embedded payload and the practical operational complexity. In particular, the need for multi-sensor calibration as well as the differences in resolutions and noise limited the reconstruction accuracy. When restricted to the use of a single wide-beam SONAR, prior assumptions needed to be considered.

In order to solve this ill-posed problem, non-linear methods have been employed using multiple point of views to bound the volume of the object. When no full rotation around the object can be done, sensor imaging models have been used (both in wide-beam and side-scan SONARs) to recover the shape from the observed intensity shading. In each of these situations, the reconstruction was obtained based on the presence of shadows and assumptions on the background (seabed). When a detailed imaging model was assumed, the a-priori lack of knowledge of the key elements such as the reflectivity of the elements, the beam pattern or geometry of the object required to either make strong assumptions or adopt a partial modelling, making the inversion ill-posed and computationally expensive. Inspired by optical reconstruction techniques, feature-based approaches have been investigated with relatively little success due to the nature of SONAR imagery, only providing sparse reconstructions.

3D reconstruction from SONAR data is therefore a large topic where multiple trade-offs between accuracy, operation time, processing time and cost of the equipment need to be made. Our work focuses on 3D reconstruction from imaging SONARs of any vertical aperture, investigating in particular the trade-off between spatial coverage and accuracy when using imaging SONARs. Importantly, the use of a single sensor of any vertical aperture enables to leverage the presence of wide-aperture imaging SONARs which are commonly embedded on ROV and AUVs for monitoring and collision avoidance and avoids excessive costs of integrating bulky and expensive additional sensors such as 3D SONARs. In this study, we assume the availibility of navigation data but do not require mm-level accuracy as typically required by SAS techniques. This enables reconstruction of data acquired by standard ROV and AUVs where navigation is based on inexpensive DVLs. Thanks to their high-frequency designs, 2D imaging SONARs typically acquire short-range (1 to 20m distance) measurements in a few tenths of milliseconds (20m-long two-way propagation takes 30ms). We consider vehicle motions of up to 20cm/s meaning that the motion during the acquisition amounts to a few millimeters only and can

therefore be ignored. In contrast to previous work, our reconstruction techniques do not assume the presence of shadows or background in the image. Our methods do not require any strong constraint on the motion of the vehicle but require a set of observation acquired along the direction of uncertainty of the sensor (elevation angle). Our methods are based on a simple imaging model and the use of optimization (deconvolution method described in chapter 5 allows modelling error mitigation) and do not rely on feature observation.

# Chapter 4

# 3D reconstruction by space carving

## 4.1 Introduction

As detailed in chapter 2, the 2D imaging SONAR model compresses 3D information on a 2D plane through an integration along the vertical direction. While this integration enables the observation of a large number of scatterers at a time, it makes the reconstruction of the initial 3D scene from a single SONAR image ill-posed. In this context, additional views of the scene and a data association technique are required to address the reconstruction problem.

In chapter 3, we provided a summary of the research work previously carried on 3D reconstruction from SONAR data. In particular when measuring bathymetry, previous reconstruction techniques from 2D imaging sensors are limited to small-aperture sensors from which range profiles can be extracted and approximated to a 3D profile. When using wide-beam sensors, state-of-the-art reconstruction techniques (Aykin and Negahdaripour [2016], Zerr and Stage [1996]) estimate the information on the last dimension by taking advantage of the acoustic shadow visible when the object is placed on a flat surface. This assumption makes the 3D reconstruction impossible when the objects are not surrounded by any flat surfaces. In addition to this, the reconstruction is obtained from a set of observations acquired by successive rotations around the object as well as a rotation of the sensor in the case of Aykin and Negahdaripour [2016]. When observing a marine

environment with an AUV, this is in general impractical either due to the uncertainty on the environment or the geometry of the scene.

In this chapter, we present a non-linear 3D reconstruction technique enabling the reconstruction of the 3D scene from a set of observations at known positions. As opposed to previously mentioned methods, our 3D reconstruction technique does not require any background to estimate the elevation information.

We first formulate the reconstruction problem as the estimation of the reflectivity from a set of integral measurements. We then present the so-called space carving reconstruction method by deriving the theoretical background based on the 2D SONAR imaging model. We introduce a three-step carving algorithm allowing online 3D reconstruction during the inspection process with little constraint on the trajectory. We assess the quality of the reconstruction on both simulated and real datasets with sensors of different vertical apertures, respectively representing a pencil-beam and a wide-beam SONAR. While assessing the quality of the reconstruction, we provide both qualitative and quantitative analysis. We finally assess the viability of the reconstruction technique for real-world objects reconstruction using both water tank and field data on multiple man-made structures including real oil field structures.

## 4.2 Formulation of the reconstruction problem

Based on the imaging model of a 2D SONAR, we formulate here the reconstruction problem. Recalling the SONAR imaging model presented in equation 2.13:

$$I_{Sonar}(r,\phi) = I_0 \int_{A_v} B_v(\theta)^2 \, S_s(r,\theta,\phi) \, \mu(r,\theta,\phi) \, \frac{\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}}{\|\vec{v}_{\theta\phi}\| \, \|\vec{n}_{r\theta\phi}\|} \, \mathrm{d}\theta$$

We adopt the following notation:

$$f(r,\phi,\theta) = S_s(r,\theta,\phi) \, \mu(r,\theta,\phi) \, \frac{\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}}{\|\vec{v}_{\theta\phi}\| \, \|\vec{n}_{r\theta\phi}\|}$$

Through this notation, we formulate the reconstruction problem as the estimation of the space occupancy at a point $P$ from a set of observations of $f$ acquired in different points

of views, each measurement at $f(r, \phi, \theta)$ being blurred by the vertical aperture $A_v$:

$$I_{Sonar}(r, \phi) = I_0 \int_{A_v} B_v(\theta)^2 \, f(r, \phi, \theta) \, \mathrm{d}\theta \tag{4.1}$$

Although $\vec{n}_{r\theta\phi}$ is *a-priori* unknown, the range of observation angles defined by the vectors $\vec{v}_{\theta\phi}$ and $\vec{n}_{r\theta\phi}$ at each point in the scene can be assumed to be small due to the limited vertical aperture of the sensor (maximum 20°). Under these conditions, the dot product $\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}$ varies little and can be considered as constant accross all observations. This assumption allows us to estimate $f$ independently from the orientation of the surface with respect to the sensor. This latter combined with the assumption of locally accurate navigation data (maximum 2cm drift per meter) provides photo-consistency. Note that the orientation-independent model is equivalent to considering the scatterers as observed locally as being small spheres of equal sizes exhibiting different albedos depending on the type of material and whether or not the space is occupied by a solid material or water:

$$\forall \, (r, \theta, \phi) \in [0, \infty[ \times [-\pi, \pi[ \times [-\pi, \pi[ \;\; S_s(r, \theta, \phi) \, \frac{\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}}{\|\vec{v}_{\theta\phi}\| \; \|\vec{n}_{r\theta\phi}\|} = C \tag{4.2}$$

with $C$ a constant value. Intuitively, this modelling is equivalent to considering the reflectivity as a first order approximation for the occupancy and is supported by the fact that we base our reconstruction method on multiple observations of the same points with little diversity in point of views thus enabling consistent estimation of $C$:

$$f(r, \phi, \theta) \approx C \, \mu(r, \theta, \phi) \tag{4.3}$$

We therefore aim at estimating the space occupancy through the estimation of the reflectivity from a set of blurred observations.

## 4.3 Space carving

In this section, a presentation of the carving reconstruction technique is given. Based on the imaging model of the SONAR, we first present the theoretical background of the carving technique. We then provide an illustration on a simple example of the carving principle.

### 4.3.1 Theory

Based on equation 4.1, one can consider subdivisions of the vertical aperture $A_v$ in $p$ uniform sections $\Theta_i$ of size $\Delta_\theta = (\theta_2 - \theta_1)/p$ such that:

$$[\theta_1, \theta_2] = \bigcup_{i=1}^{p} \Theta_i = [\theta_1, \theta_1 + \Delta_\theta] \cup \cdots \cup [\theta_1 + (p-1)\Delta_\theta, \theta_2]$$

and

$$I_{Sonar}(r, \phi) \quad = \quad \sum_{i=1}^{p} \int_{\Theta_i} B_v(\theta)^2 \;\; f(r, \theta, \phi) \; d\theta = \sum_{i=1}^{p} \varsigma_i(r, \phi) \tag{4.4}$$

with

$$\sigma_i(r, \phi) = \int_{\Theta_i} B_v(\theta)^2 \;\; f(r, \theta, \phi)$$

Due to the positivity of $f$, each subsection integral $\varsigma_i(r, \phi)$ is positive and for small enough subdivision $\Theta_i$, one can assume (Riemann quadrature):

$$\varsigma_i(r, \phi) \approx K_{\Theta_i} \; B_v(\theta_i)^2 \; f(r, \theta_i, \phi)$$

with $\theta_i$ being $\Theta_i$ midpoint and $K_{\Theta_i}$ being a measure of the size of the integration domain $\Theta_i$. For simplicity and since we aim at estimating voxels of equal sizes, we consider, without loss of generality $K_{\Theta_i}$ as a constant equal to 1.

From this approximation, results the following inequalities:

$$\forall i \in [\![1, p]\!]; \quad 0 \leq f(r, \theta_i, \phi) \leq \frac{I_{Sonar}(r, \phi)}{B_v(\theta_i)^2} \tag{4.5}$$

Each observation provides an upper limit to the intensity reflected by the scattering points within the vertical aperture of the sensor and in particular in the case of $N$ observations $\{I_k\}_{1 \leq k \leq N}$ of the same point $P$ from different elevation angles $\{\theta_k\}_{1 \leq k \leq N}$, we have:

$$0 \leq f(P) \leq \min_{1 \leq k \leq N} \frac{I_k}{B_v(\theta_i)^2} \tag{4.6}$$

An estimated upper bound to $f(P)$ is therefore given by the observation of a minimum scaled intensity. The estimation of this upper bound from the set of observations

$\{I_k\}_{1\leq k\leq N}$ and their associated vertical angles of observation $\theta_i$ constitutes a reconstruction method. Since this non-linear approach relies on the observation of empty spaces to bound the function $f$, we refer to it as space carving.

### 4.3.2 Illustration of the carving reconstruction principle

In 4.1, an illustration of the carving reconstruction technique is provided. As shown in figures 4.1-a and 4.1-b, the acquisition process is performed along the direction of uncertainty caused by the vertical aperture of the 2D multibeam sensor. This direction is perpendicular to the image plane of the sensor at each measurement and referred to as the U-axis. Note that this axis is represented here as a vertical axis but could in practise correspond to any 3D line as it only depends on the orientation of the sensor. The scene is observed from 4 locations at different altitudes under an horizontal angle $\phi$ and at a range $r$. The overlap between the footprints enable multiple observations of the points located on the U-axis. Figure 4.1-c exhibits the multiple intensities measured when moving along the Z-axis while figure 4.1-d presents the reconstructed axis using the carving principle.

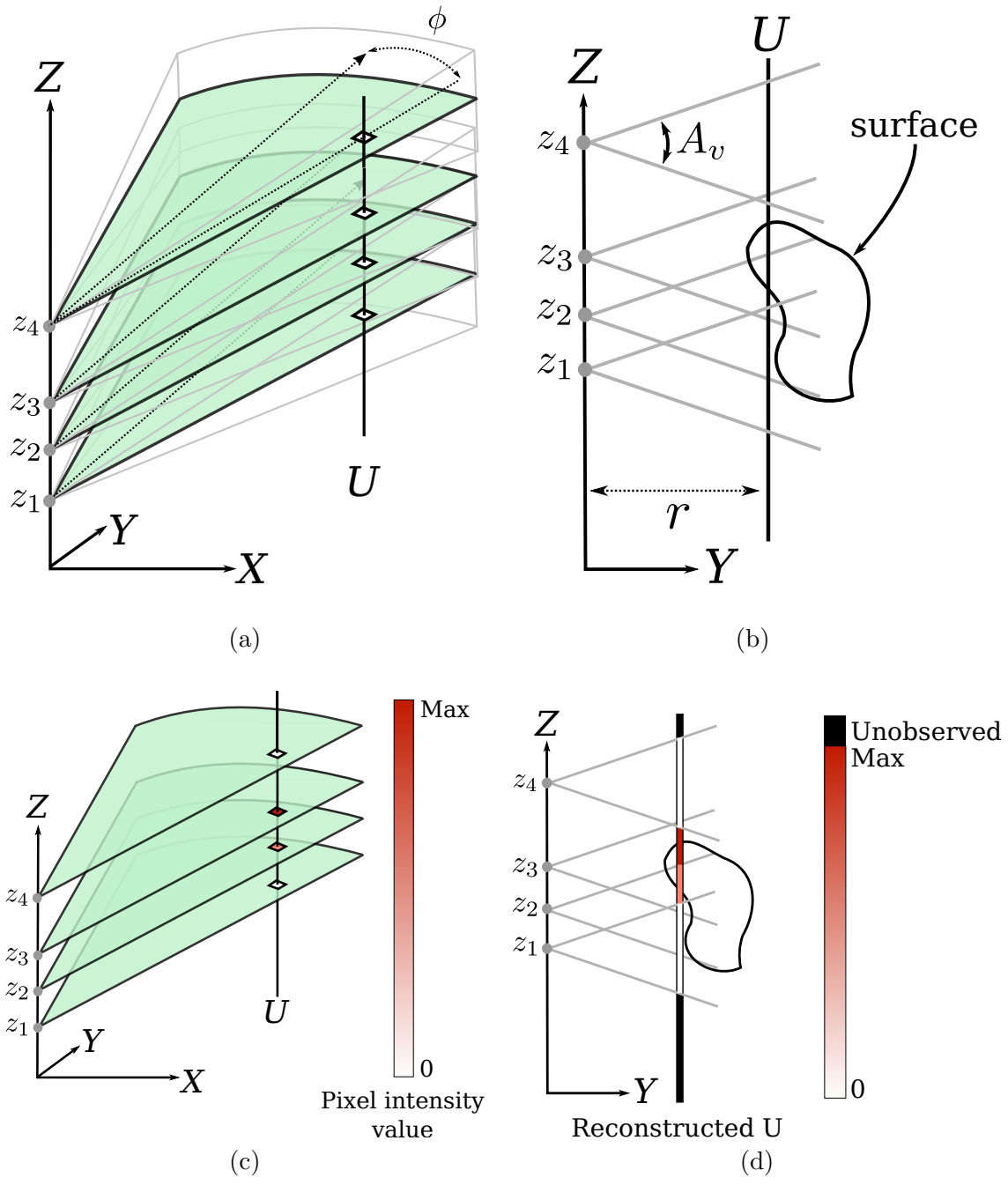



FIGURE 4.1: Carving reconstruction principle.
a) Geometry of samples acquisition along the direction of uncertainty of the sensor: U-axis. 4 measurements are made at different vertical positions $z_1$, $z_2$ and $z_3$. b) The samples provide multiple observations of the surface of the object. Due to the vertical aperture of the sensor $A_v$, the spatial extent of the measurements exhibit overlapping areas. c) Depending on the presence of scatterers in the aperture of the sensor, the samples feature different intensities ranging from low to high intensities. d) The reconstruction of the observations along the axis U is obtained by keeping the minimum observed intensity for each vertical section. The observation of empty spaces (white intensity) provide a spatial boundary to the object.

## 4.4 Space carving algorithm

In addition to the principle presented in the previous section and due to the nature of SONAR imaging, several additional processing steps are needed. We present here the multiple steps of the practical algorithm implementing the space carving technique.

### 4.4.1 Overview

As illustrated in figure 4.2, the carving reconstruction algorithm is composed of three major steps. The first two steps aim at generating a temporary map of the environment

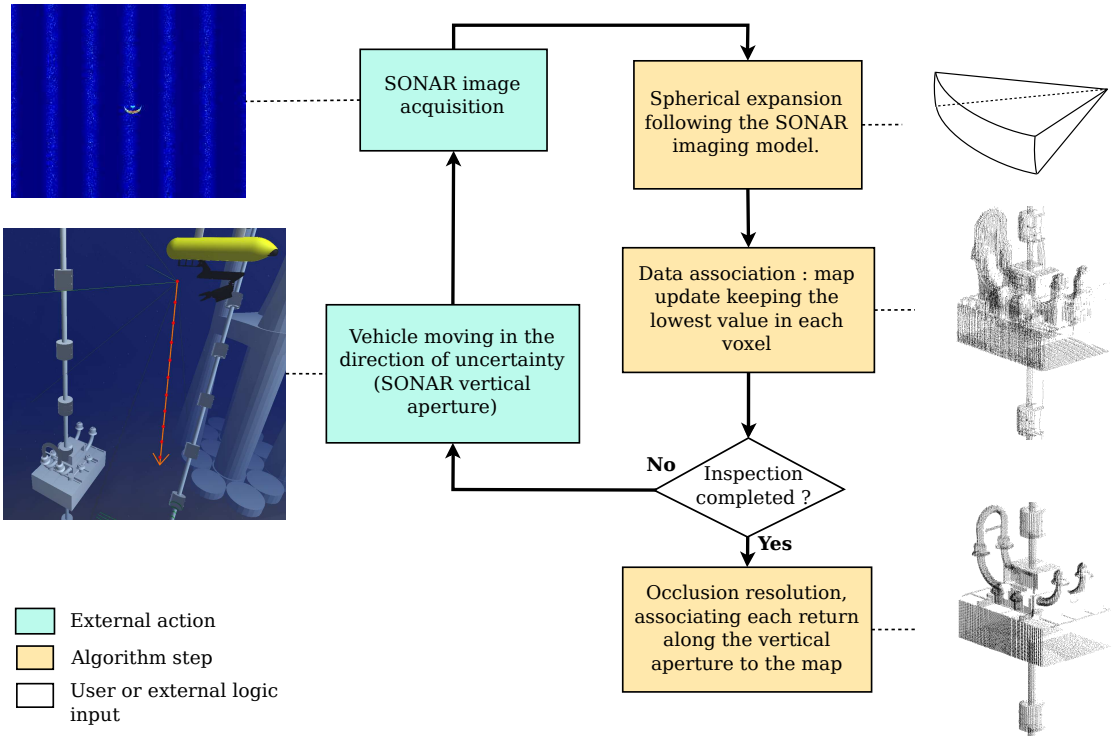



FIGURE 4.2: Carving algorithm diagram. The 3-step carving algorithm features two initial steps applied each time new data is acquired. While these two steps provide an initial temporary estimate of the occupancy, the occlusions happening during the imaging process are not solved at this point. Once enough data has been acquired, a final occlusion resolution step provides the final map.

and are applied each time new data is acquired. Based on the SONAR imaging model, a spherical expansion of the 2D SONAR image to a 3D discrete representation is first performed. Then the carving filtering rule is applied by comparing the new 3D data to the current estimate of the map and keeping the lowest observed intensity. New measurements are then acquired at various positions along the direction of uncertainty

(vertical aperture) to refine the map until the inspection is judged to be completed. Once this condition satisfied, a last step called occlusion resolution is applied to generate the final map, taking into account the occlusions happening during the imaging process.

### 4.4.2 Spherical reprojection principle

The SONAR imaging process compresses 3D information into a 2D representation by summation along the vertical aperture. In order to retrieve a 3D representation of the scene, the 2D images are first "expanded" in a spherical way using a 3D rotation. As illustrated in figure 4.3, for each return at a bearing angle $\phi$, the ensemble of possible scatterer position is obtained by 3D rotation around the vector $\vec{v}_\phi$ of a rotation angle within the vertical aperture $[-\frac{A_v}{2}, \frac{A_v}{2}]$. The rotation axis vector $\vec{v}_\phi$ is defined by the cross product between the bearing direction $\vec{u}_\phi$ and the unit vector orthogonal to the sonar plane $\vec{z}_{Sonar}$.

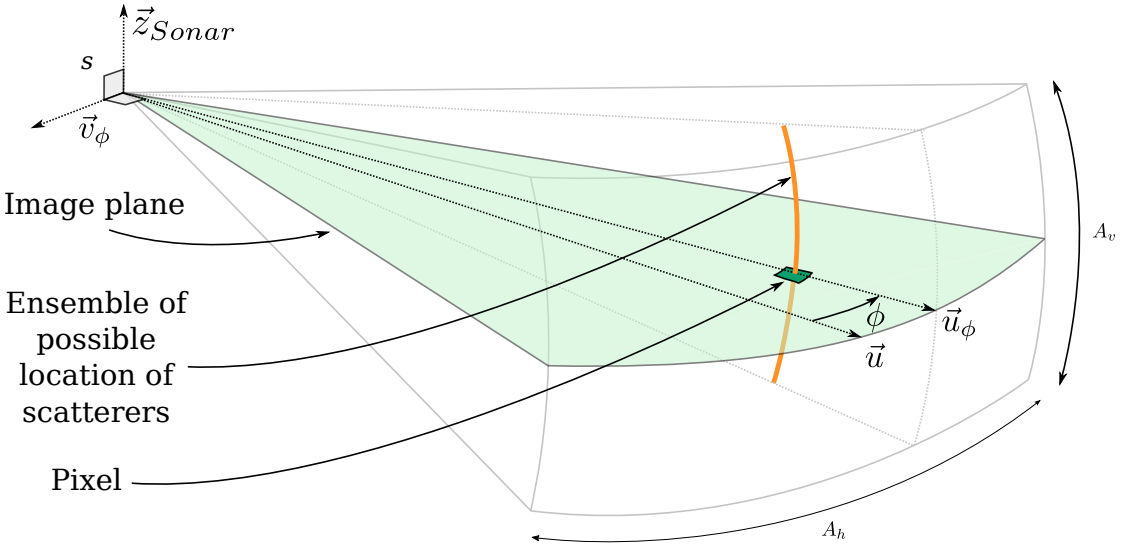



FIGURE 4.3: Spherical reprojection. For each pixel, the ensemble of possible scatterers related to it is obtained by applying a 3D rotation of the pixel centre around the axis described by the vector $\vec{v}_\phi$.

Once the spatial extent of potential scatterers defined, the resulting spherical arc is then discretized at a given angular resolution $\theta_{res}$. As visible in figure 4.4, the discretization step provides multiple 3D points for each pixel.

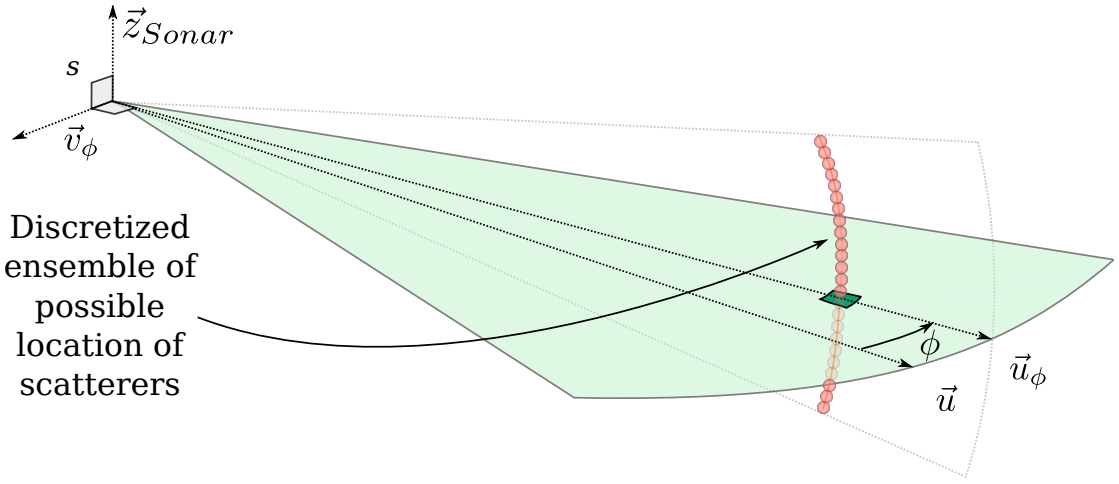



FIGURE 4.4: Scatterer candidates. The ensemble of possible scatterer position for the pixel is discretized at a given angular resolution, providing a set of 3D points for each pixel.

Importantly, an intensity is assigned to each generated 3D point based on its elevation angle following the upper bound described in equation 4.5:

$$I(r, \phi, \theta) = \frac{I_{Pixel}(r, \phi)}{B_v(\theta_i)^2} \tag{4.7}$$

This representation enables direct 3D filtering by maintaining at each time the estimated upper upper bound for the observed intensity at a 3D location.

### 4.4.3 Occlusions and conservative reprojection

As described in section 2.2.5, the absence of acoustic return at a given point can either be the result of open-water propagation of the emitted wave or a shadowing effect due to the presence of an object in between the sensor and the point considered. As result and when operating a spherical reprojection of a zero intensity pixel, it is necessary to check for potential occlusion. In this situation, a conservative reprojection where zero intensity pixels are projected only when no non-zero intensity pixel is present in between the considered pixel and the sensor is needed.

Figure 4.5 provides an illustration of the conservative reprojection rule where the pixels measured along the directions $\vec{u}_1$ and $\vec{u}_2$ are expanded to 3D. In the first case, no acoustic return has been measured along the $\vec{u}_1$ direction, all intensities can therefore be reprojected. In the second case, the axis directed by $\vec{u}_2$ features two acoustic returns

which can be both reprojected. In the last situation, an empty pixel is measured behind an acoustic return, since it could be due to both an absence of scatterer or an occlusion by a scatterer at shorter range (the scatterer leading to the return at shorter range), no reprojection is operated.

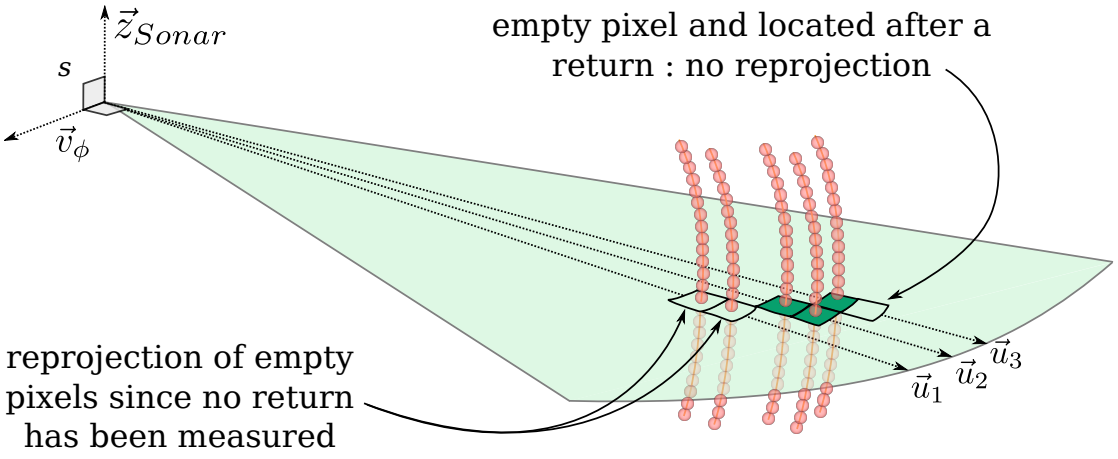



FIGURE 4.5: Conservative reprojection. In order to account for potential occlusions when expanding a SONAR image to 3D, empty pixels are only reprojected if no acoustic return has been measured between its position and the sensor. The pixels measured along the directions $\vec{u}_1$ and $\vec{u}_2$ are therefore reprojected while the empty measurement on $\vec{u}_3$ is not expanded due to a full pixel measured at shorter range in this direction.

This conservative reprojection rule enables to obtain a set of hypothetical locations for scatterers in 3D. At this stage, each 3D point generated represents a possible source of backscattering with an associated intensity representing its *a-priori* maximum backscattering strength or reflectivity.

### 4.4.4 3D data storage and Min-filtering

Once a 3D representation obtained, a 3D map gathering the multiple observations obtained is maintained. In order to enable search and general processing operations, a structured representation is needed. 3D data in general requires large amounts of memory making a full regular gridded cube inconvenient. For this reason, an octree structure (Meagher [1982]) is chosen to provide a lightweight representation. An Octree is a k-d tree structure with nodes splitting in 3D, producing 8 partitions. It is constructed by recursively subdividing space into eight cells until a limit condition is met. This condition can either be a pre-defined number of elements or a maximum tree depth. As illustrated

in figure 4.6, the possibility not to allocate certain child nodes allows sparsity of the data and decreases the memory usage in comparison to regular grids.

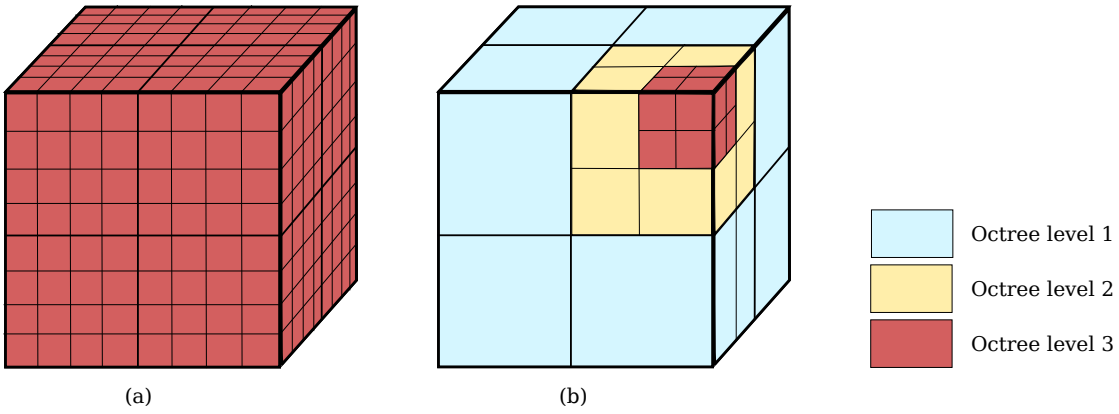



FIGURE 4.6: Octree structure. Compared to a regular grid (a), an octree structure (b) enables a lightweight memory usage by decreasing the number of levels required to store unoccupied cells.

Once a new SONAR image has been expanded to a 3D representation, the set of 3D points is added to the map. When adding a point to the map, if the voxel the new point fall into had previously been populated by another measurement, the carving method is applied: the intensity associated to the considered voxel is the lowest estimated reflectivity of the two measurements. Thus if the new reflectivity measurement appears to be lower, its value will be affected to the voxel, otherwise, the previous measurement will be kept. If the new measurement appears to fall into a previously observed voxel, the intensity of the voxel is initialised to the first measurement.

Each time a new measurement is added, the map representing the best current estimate of the reflectivity is updated. Once enough samples acquired (end of inspection), this temporary map is processed to solve the occlusions happening during the inspection.

### 4.4.5 Occlusion resolution

As illustrated in figure 4.5, when multiple returns are measured on the same bearing axis (as visible on the axis directed by $\vec{u}_2$), each return is reprojected along the entire vertical aperture. This situation reflects the *a-priori* possibility of presence of scatterers at each elevation angle for each return. Due to the occlusion phenomenon described in section 2.2.5, it is impossible to observe from the same point of view two consecutive scatterers

aligned with the SONAR location. Therefore, following the conservative reprojection of the samples and their association in a map, there is a need for handling the potential occlusions that happened during the imaging process by explicitly associating each pixel to a set of scatterers located along the vertical aperture. The occlusion resolution step aims at generating a final map of the scene by only retaining the observed surface of the objects (front scatterers).

Once the initial map obtained, it is possible to solve the occlusions occurring at each point of view by comparing the SONAR images to the map. As illustrated in figure 5.29, associating each pixel to a set of 3D points along the vertical aperture enables to label points in the map as observed or shadowed. Iterating from short to long range along each bearing direction, the vertical aperture is divided in a set of intervals labelled in two categories: free or occluded. Based on this labelling, the points in the map falling into the footprint of the pixels can be labelled as observed or shadowed by a previous scatterer along the propagation direction. As illustrated in figure 4.8 the points identified

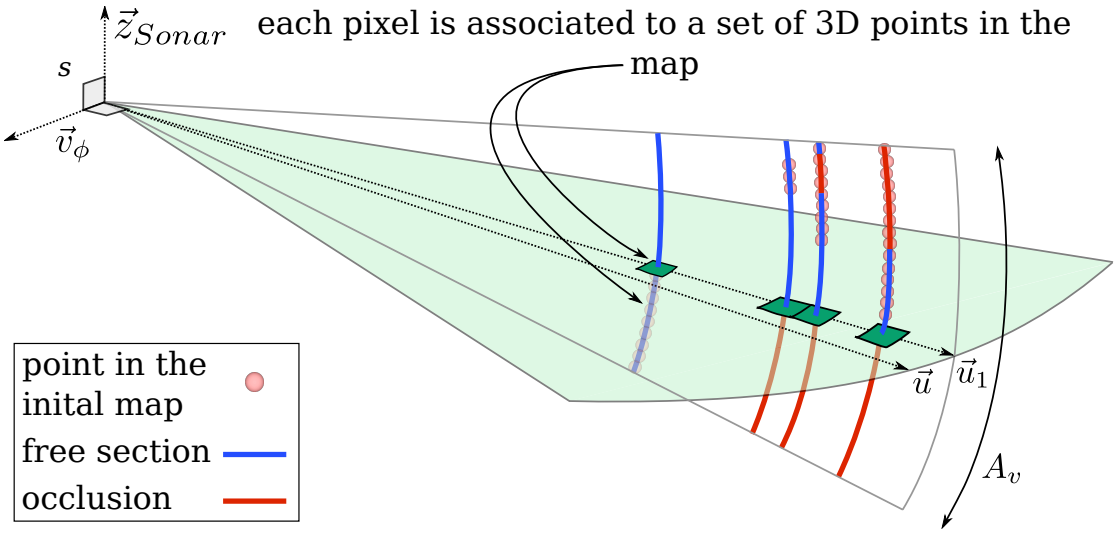



FIGURE 4.7: Occlusion resolution method. Each image is reprojected in 3D and compared to the map previously built. Iterating from the lowest range to the maximum range to the SONAR location along the $\vec{u}_1$ direction, each pixel is projected and associated to an interval or a set of intervals in the vertical aperture. Once these intervals labelled as free or occluded, the points in the map are labelled as observed or shadowed.

as shadowed or observed in the image, only the points observed are kept and added to the final map, effectively retaining the front surface points of the object. Since some points can appear in multiple images, a voxel-based filtering is applied when adding each

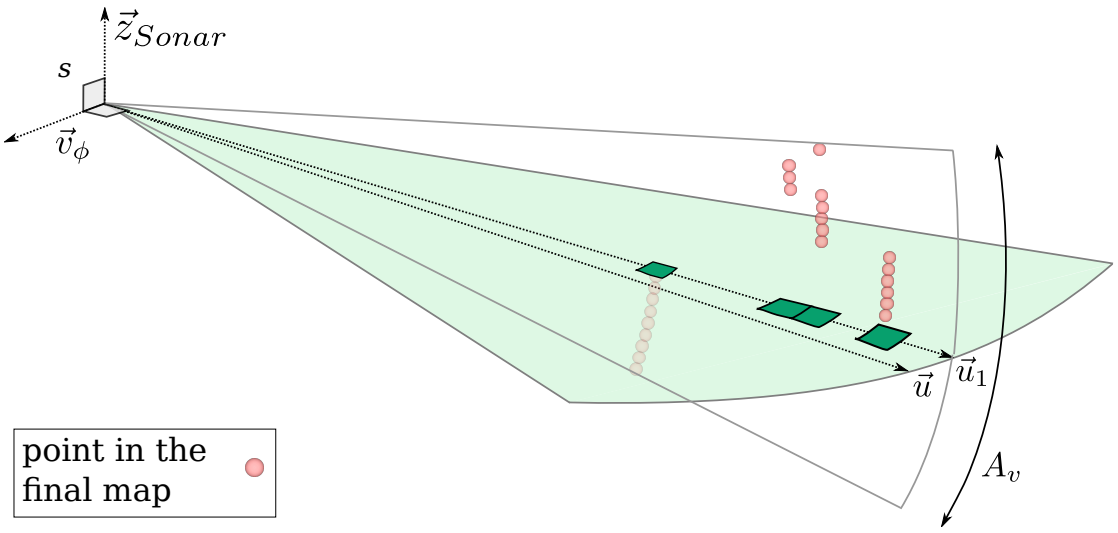



FIGURE 4.8: Solved occlusions. Once the points in the map labelled, only the points observed in the SONAR image considered are retained and added to the final map.

new set of points to the final map, keeping one point per voxel. An intensity value is kept for each point, keeping once again the lowest reflectivity observed (carving rule).

### 4.4.6 Non-uniform and non-regular sampling

Due to the imaging SONAR geometry model (see figure 4.9), each measurement provides samples spaced at regular $\Delta_r$ range intervals and $\Delta_\phi$ horizontal aperture angular intervals leading to a non-uniform and non-regular sampling in the cartesian space. Similarly, the sampling period in the remaining dimension (along the vertical aperture) is assumed to be performed at any sampling period $\Delta_z$. While the samples density is anisotropic, octree structures represent a regularly sampled grid. In order to optimize the space usage, samples are stored at cartesian resolutions $(\Delta_u, \Delta_v, \Delta_w)$ as described in figure 4.10. In order to adapt to the spherical samples to the uniform representation of an octree, a scaling is applied in every direction making the resolutions in all dimensions equal to the minimum resolution $\Delta_o = min(\Delta_u, \Delta_v, \Delta_w)$. The points are then stored in an octree enabling the carving processing at $\Delta_o$ resolution. When generating the final map, an inverse scaling is applied on the map to restore the anisotropic resolutions on (u,v,w) grid.

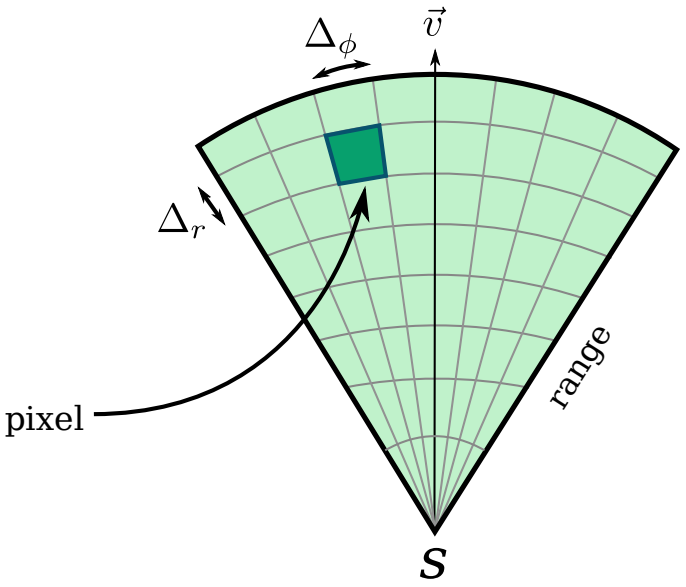



FIGURE 4.9: Non-uniform sampling. Due to the SONAR imaging geometry, the cartesian space XYZ is sampled regularly in spherical coordinates at regular $\Delta_r$ range intervals and $\Delta_\phi$ horizontal aperture angular intervals.

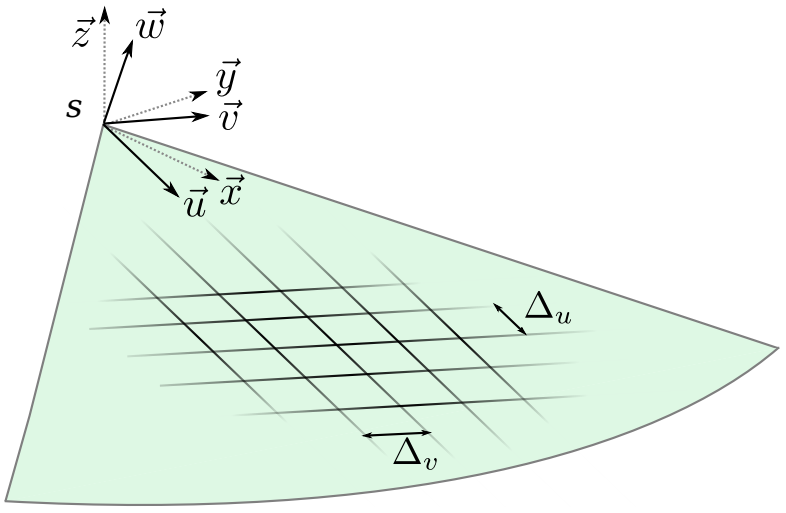



FIGURE 4.10: Cartesian gridding. In order to adapt to the regular sampling of an Octree, a cartesian grid based on the first sample is used.

### 4.4.7 Image denoising and intensity threshold

When reconstructing from noisy images, the observation of empty spaces is made difficult by noisy measurements. In order to enable the distinction between noise and data, a denoising process can be applied in two ways.

**Statistical denoising of the input samples**

As described in section 2.5.2, a noise model can be obtained from a set of open-water observations by estimating the first two moments of the noise distribution at each location in the image $(r, \phi)$. The denoising criteria being scaled on the local noise characteristics $(m, \sigma)$, this method is adapted to the non-uniformity of typical SONAR noise patterns. Once the noise distribution known, a probability of the pixel to represent noise can be computed using the Z-test value:

$$Z_{Score}(r, \phi) \; = \; \frac{I(r, \phi) - m}{\sigma} \; > \; T \tag{4.8}$$

with T being a user defined threshold. As illustrated in table 4.1, applying a threshold typically chosen between 2 and $3\sigma$ to a gaussian noise distribution enables to suppress most of the noise measurements.

| T | % of noise points removed |
|---|---|
| 1 $\sigma$ | 84.1 |
| 2 $\sigma$ | 97.7 |
| 3 $\sigma$ | 99.9 |

TABLE 4.1: Proportion of points removed by a Z-test of threshold T on a gaussian noise distribution: percentage of points equivalent to the complementary cumulative distribution function of a gaussian distribution evaluated at respectively $(\sigma, 2\sigma, 3\sigma)$.

**Static threshold**

Once the temporary map generated, a set of 3D points with associated intensities is obtained. At this stage, a static threshold can be applied to remove the points of lowest intensity without needing any prior knowledge on the noise model.

### 4.4.8 Implementation

The carving algorithm has been implemented in C++ to allow for high execution speed, object-oriented programming with the possibility to build on a large range of open-source libraries, providing interfaces to various hardware platforms as well as many standard

data types and protocols. An implementation in C++ therefore makes easier a future integration on an underwater vehicle for field data validation trials. Our implementation took advantage of the OpenCV library (Bradski et al. [2000]) for fast 2D image processing and data structures. Similarly, our implementation relies on the Point Cloud Library (Rusu and Cousins [2011]), commonly used for 3D processing applications such as object recognition, 3D reconstruction, SLAM, etc. In particular, PCL offers many 3D data-related data types and structures as well as multiple implementations of the Octree structure. As presented in section 2.5.1, our sonar simulation framework is based on ROS middleware as well as a library developed in the Heriot Watt Ocean System Laboratory implementing multiple basic functionalities of an underwater vehicle such as point to point navigation, thruster control and human interfaces to add functionalities to the UWSim framework.

## 4.5 Simulation results

We present here simulation results obtained with the UWSim environment. First, the reference models used in the simulation environments are presented. Following this and in order to validate the concept of carving method and occlusion resolution, simulated experiments are presented with noise-free images. Both qualitative and quantitative analysis are presented. A short study of the influence of noise is then presented in order to evaluate the suitability to the reconstruction method to real data.

### 4.5.1 Reference models used in simulation

In order to evaluate the reconstruction method, four reference models were used allowing both qualitative and quantitative analysis on objects of different shapes and sizes. As illustrated in figure 4.11, the reference models represent medium to large man-made objects similar to typical underwater man-made objects and exhibit different levels of details. The first object is a simple sphere featuring a full range of angles of incidence to the surface, similar to a buoy. The second object is a cylinder equipped with a wheel valve as could be found on a oil field. The third object typically represents a pillar as could be found supporting a bridge or a heavy man-made structure. The last reference model

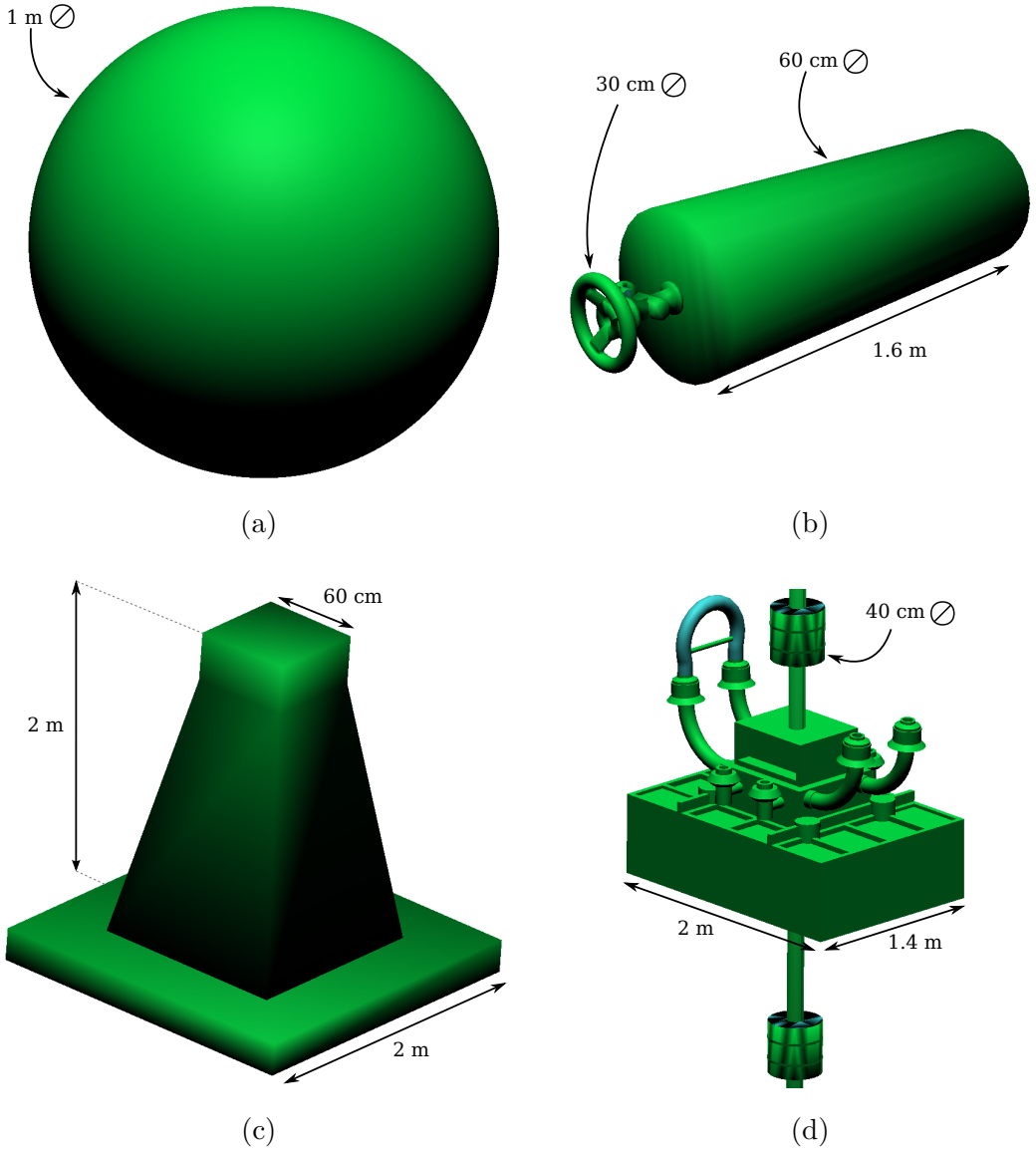



FIGURE 4.11: Reference CAD models used in simulation. a) 1m diameter sphere. b) 50cm diameter cylindrical tank with a wheel valve. c) Pillar model with a 2m wide base and a 60cm wide top. d) 2m large oil-field structure mounted on a vertical axis exhibiting 40cm diameter buoyancies.

represents a typical oil-field structure, exhibiting more details and featuring cylindrical buoyancies as frequently found on underwater man-made floating structures.

### 4.5.2 Noise-free simulation results

In order to simulate both a pencil-beam imaging SONAR and a wide-aperture SONAR, experiments were made with both a 1° and a 10° vertical aperture simulated SONAR. The sensor featured a range resolution of 2cm and a bearing resolution of 0.2°. Figure

4.12 presents reconstruction results obtained with respectively a pencil-beam sensor (b,e,h,k) and wide-beam sensor (c,f,i,l).

#### 4.5.2.1 Quantitative analysis metrics

In order to quantify the accuracy of the 3D reconstructions, a study of the error (distance of each reconstructed point to the CAD model) distribution is necessary. Figure 4.13 depicts a typical distribution of the reconstruction error, obtained from the unsigned reconstruction errors of the four models presented in section 4.5.1 at a fixed vertical sampling period of 1cm and a 10° vertical aperture SONAR. As visible in the figure, the distribution exhibits a large tail due to the presence of outliers. When using noise-corrupted images as in real data, the presence of outliers is typically expected to increase making necessary the use of a set of robust metrics to study the reconstruction error.

As a consequence, four metrics were chosen to compare the final point clouds to the original CAD models:

a) Unsigned median error: as well known in statistical analysis, the mean estimator of a heavy-tailed distribution leads to artificially high values due to the exaggerated weight of the outliers. In this situation, the mean estimate reflects more the presence of outliers (small part of the population) than the typical values as can be observed in a large part of the population. In comparison to the mean error, the median value offers more robustness to the estimate of the statistical error in presence of outliers by considering the error value of the most typical point of the distribution: its central point.

b) Surface coverage: while the distance of the reconstructed points to the CAD models is a measure of the precision of the reconstruction, figure 4.12 illustrates the differences in surface coverage of the inspected object: the reconstruction obtained using a pencil-beam sensor typically leads to good reconstruction accuracy (4.12-e) but low overall surface coverage compared to a wide-beam sensor (4.12-f). When reviewing the quality of a 3D reconstruction, there is therefore a need to take into account the proportion of reconstructed surface. A good estimate of the coverage is given by computing the number of points on the surface of the model for which a reconstructed point can be found within a given radius $r_c$. We typically chose $r_c = 3$ cm. The

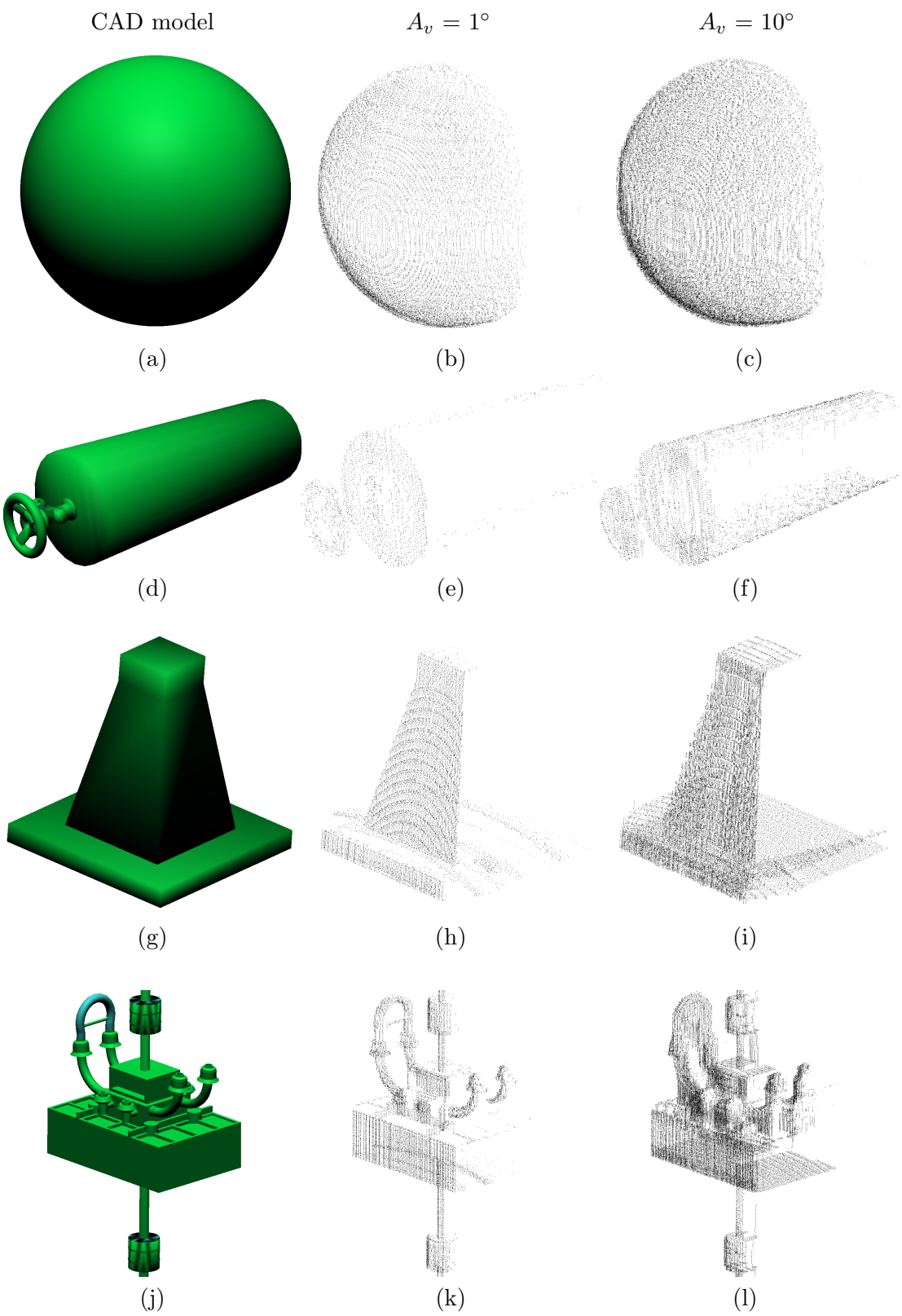



FIGURE 4.12: 3D carving reconstruction from noise-free simulated data obtained after inspection of the four reference models. Reconstructions of the CAD models (a,d,g,j) using the pencil-beam sensor (b,e,h,k) and a wide-beam sensor (c,f,i,l).

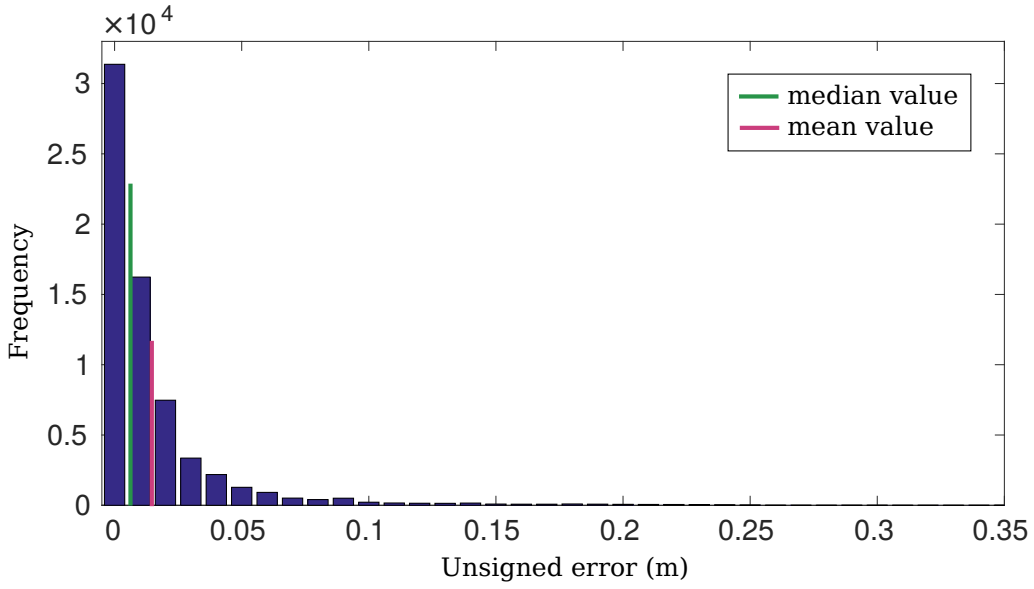



FIGURE 4.13: Typical reconstruction error distribution obtained from averaging the error distributions obtained with the four reference models with a 1cm vertical sampling period and a 10° vertical aperture SONAR. The distribution exhibits a large tail resulting in a mean value higher than its median value.

estimate of the surface coverage is then typically expressed as a percentage of the model surface.

c) Proportion of outliers: when using wide-aperture sensors and *a-fortiori* in noisy environments, the presence of reconstructed points far from the original model is commonly observed. In order to quantify the presence of these outliers, a threshold radius $r_o$ is set. Any reconstructed point being further than $r_o$ from the CAD model is considered as an outlier, the overall metric being the proportion of points satisfying the outlier condition.

d) Unsigned median error to coverage ratio: since the first two metrics, median error and coverage often appear to be traded for each other when choosing a sensing modality, there is interest in using a combined metric providing a way to compare sensors of different footprint sizes.

In all our experiments, CAD models constitute our ground truth for both qualitative and quantitive analysis. Their accuracy, both in terms of geometry and position in the scene is therefore of prime importance. When considering simulated data, the positions of the structures are defined by the user in UWSim. Similarly, the sensor offset and the position of the vehicle is known since provided by the simulator. The structures observed are

observed through direct raytracing of the CAD models therefore both the geometry of the structures and their position are known in advance and can be directly compared to the reconstructed point clouds without need for registration. When considering real data acquired in water tank, no quantitative analysis was performed due to the absence of accurate 3D CAD models. On the contrary, field experiments (described in section 4.7.2) were performed on custom-made structures for which 3D CAD models were generated (provided by the manufacturer) at a sub-centimeter accuracy. These structures were cleaned prior to deployment on the field of experiments to avoid any alteration of the geometry by marine growth and sand accumulation. Quantitative results obtained on this dataset (see table 6.1) were then obtained by co-registration of multiple point clouds (see section 6.2.2.1) and registration of the CAD models to the point clouds using the methods described in section 6.3.2.5. In both cases the results were carefully checked and provided the same level of accuracy as manually registered CAD-models.

#### 4.5.2.2 Quantitative results

We present here a quantitative analysis of the carving reconstruction technique using the four metrics presented in section 4.5.2.1. As illustrated in figure 4.14, the reconstruction median error with a wide-beam sensor is roughly twice as high as when using a pencil-beam sensor. When considering large sampling periods, the difference in error increases due to constant error values when using a pencil-beam sensor.

In addition to a 10° vs 1°comparisons, experiments were made with sensors of wider aperture (respectively 20°and 30°). Figure 4.15 shows the increase in median reconstruction error when considering larger vertical aperture sensors.

As shown in figure 4.16, the surface coverage decreases when considering larger sampling periods with values roughly twice as high when using a wide-beam sensor except on the sphere structure.

Figure 4.17 shows the increase in surface coverage when considering sensors of larger vertical apertures.

As can be observed in figure 4.18, when combining the two metrics, similar values are obtained at low sampling periods (up to 5cm). Larger differences are observed at higher sampling periods.

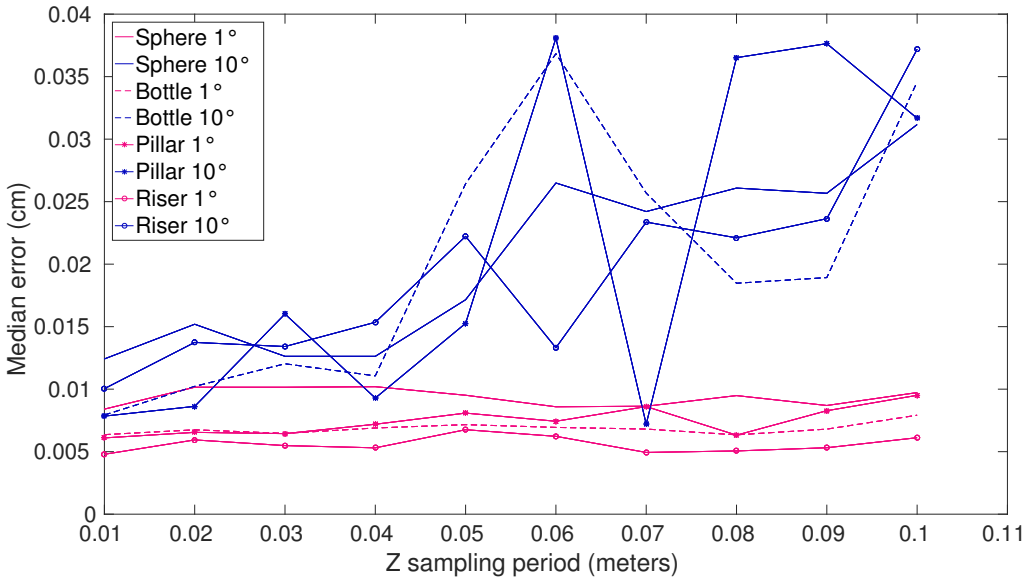



FIGURE 4.14: Median error versus vertical sampling period average on all models for different apertures with two different sensors: pencil-beam (magenta) and wide-beam (blue). For small sampling periods, the reconstruction error of the 10° aperture sensor is roughly twice as high as when using a 1° sensor.

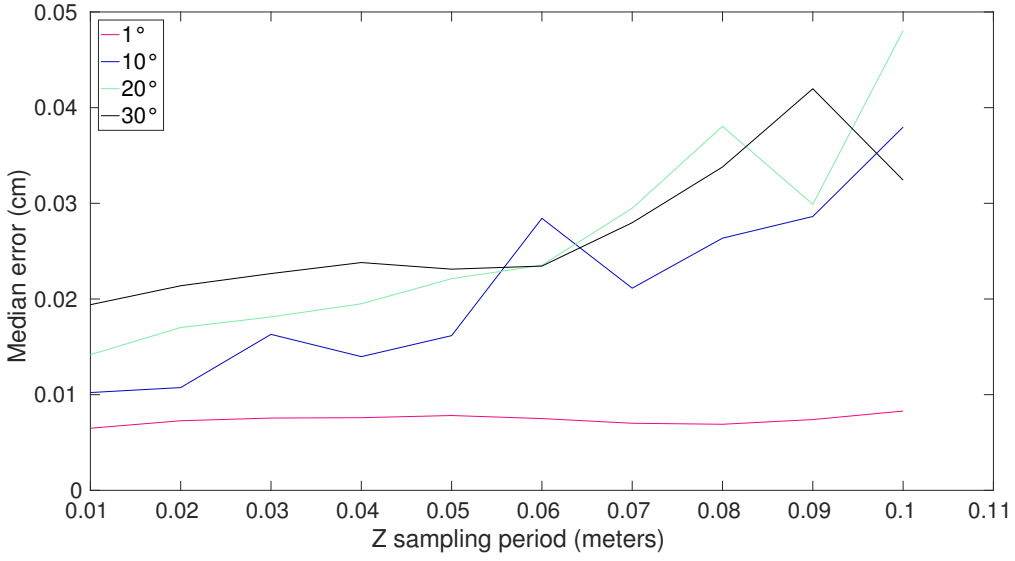



FIGURE 4.15: Median error versus averaged on CAD all models vertical sampling period for different sensor apertures.

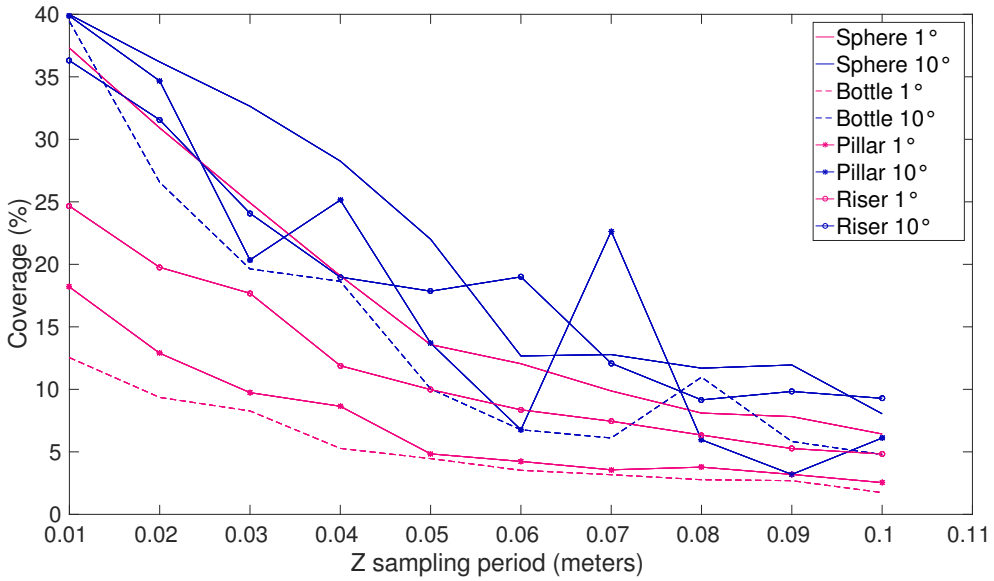



FIGURE 4.16: Surface coverage versus vertical sampling period. The surface coverage yielded when using a pencil-beam sensor is twice as low as when using a wide-beam sensor except on the sphere object where a similar coverages are observed.

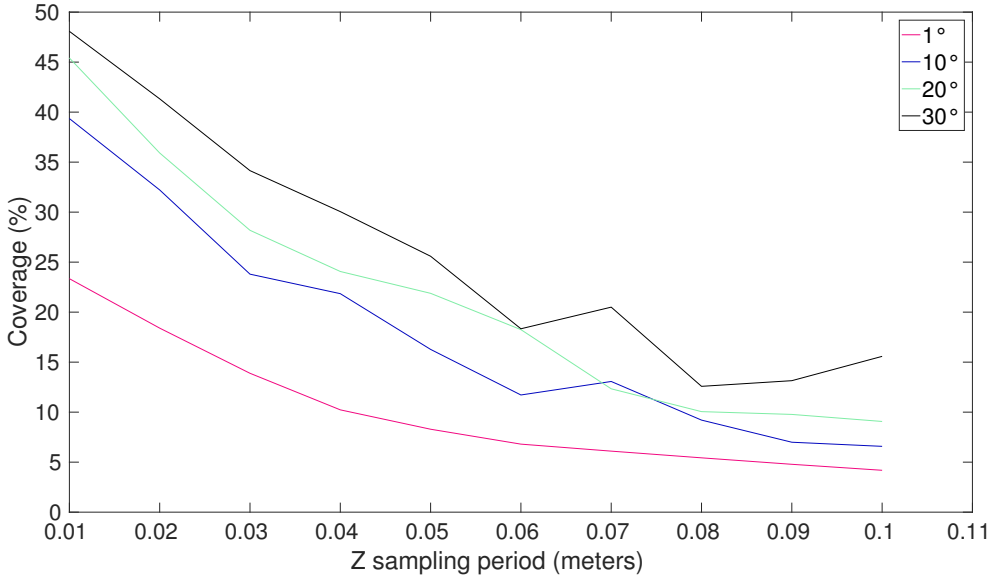



FIGURE 4.17: Surface coverage averaged on all CAD models versus vertical sampling period for different sensor apertures. Increasing the aperture consistently increases the surface coverage of the reconstructions.

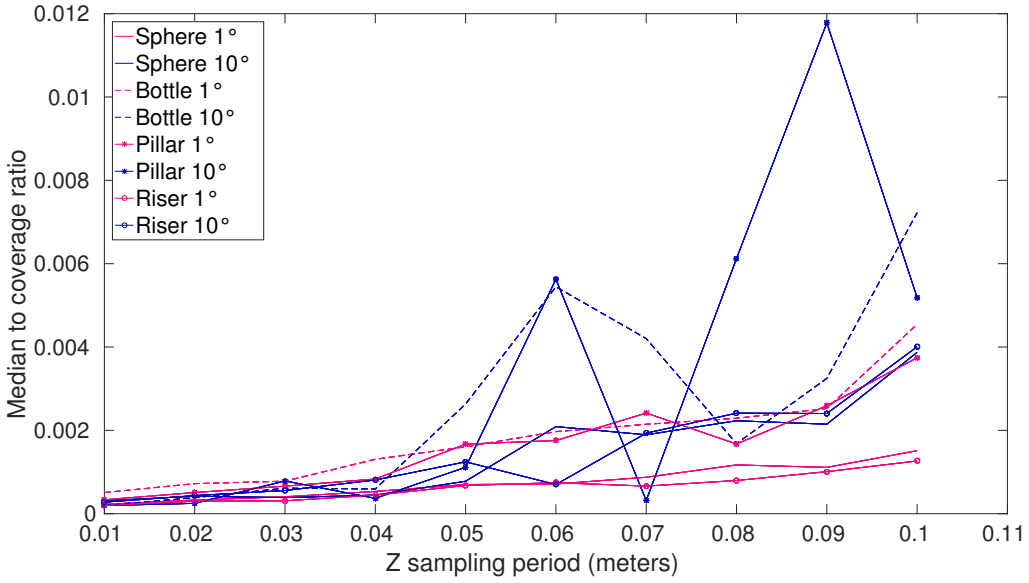



FIGURE 4.18: Median error to coverage ratio versus vertical sampling period. Similar values are observed when at low sampling periods (up to 5cm) with larger variations when using large sampling periods.

As illustrated in figure 4.19, the median error to coverage ratio of reconstructions obtained with various apertures are very similar at low sampling periods.

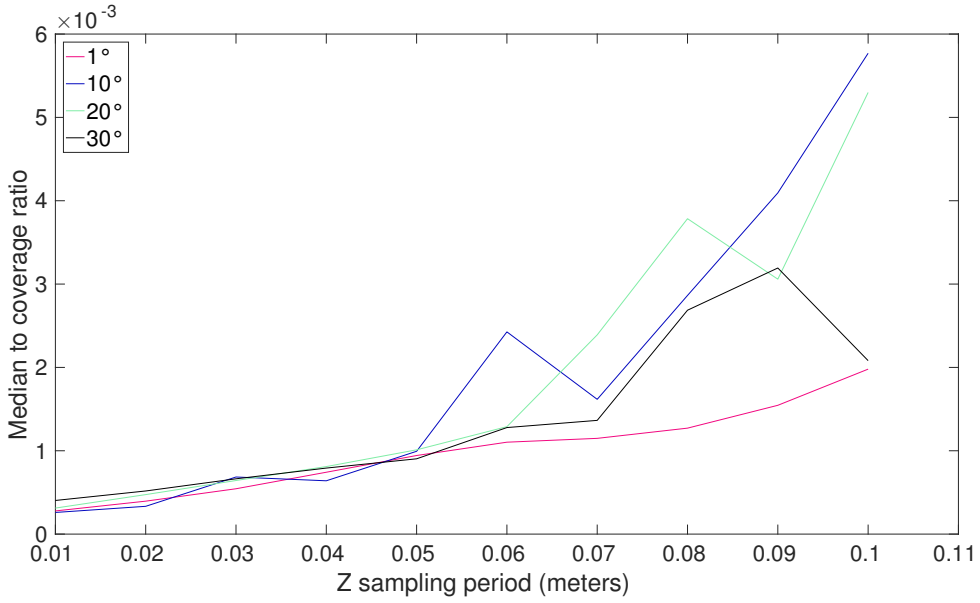



FIGURE 4.19: Median error to coverage ratio averaged on all CAD models versus vertical sampling period for different sensor vertical apertures.

When studying the presence of outliers (fig.4.20), it can be observed that the use of a

pencil-beam sensor leads to less than 1% of outliers. Reconstructions from wide-beam sensors contain larger proportions of outliers but remain lower than 15% (at most) when using small sampling periods.

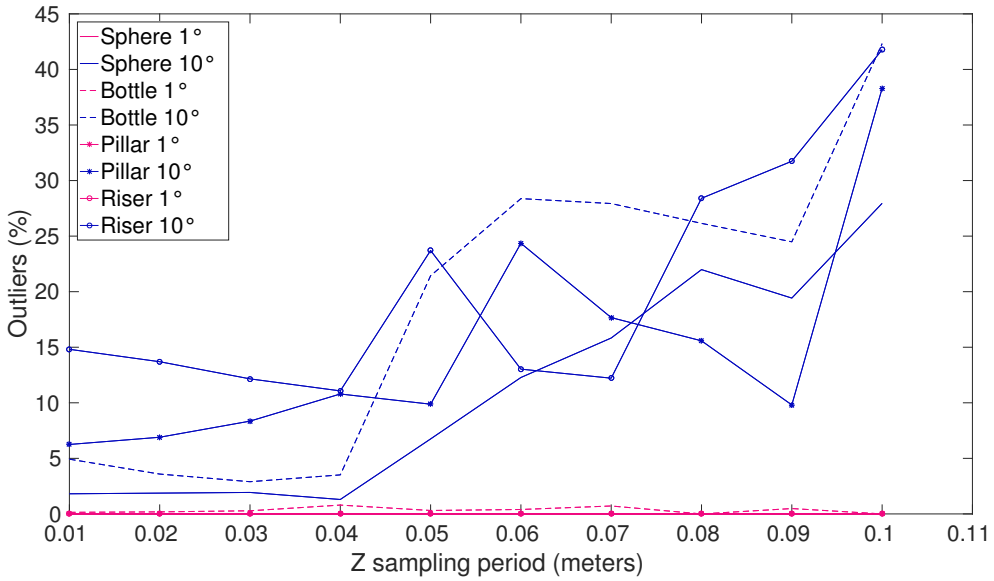



FIGURE 4.20: Presence of outliers versus vertical sampling period using a pencil-beam sensor and a wide-beam sensor.

#### 4.5.2.3 Analysis

On a qualitative level, the reconstructions from noise-free data presented in figure 4.12 showed the difference between a wide and a pencil-beam sensor. The pencil-beam reconstructions (b,e,h,k) exhibit good resolution but only partial reconstructions: in particular, figures (e,h,k) show the impossibility of observing horizontal surfaces. In opposition to this, the wide-beam reconstructions (c,f,i,l) offer increased coverage at the cost of lower geometrical details.

As visible in figures 4.14 and 4.16, opposite behaviours of the two first metrics are observed when using various sampling periods resulting in similar median to coverage ratio values (see figure 4.18). While the use of a wide-beam sensor provides higher surface coverage by scouring large amounts of water at a time, the width of the aperture induces a blurring in the observations making the recovery of the initial 3D geometry more complicated. This result therefore shows the interest of the carving reconstruction

method where overall similar results are obtained with a wide-aperture SONAR as when using a pencil-beam sensor.

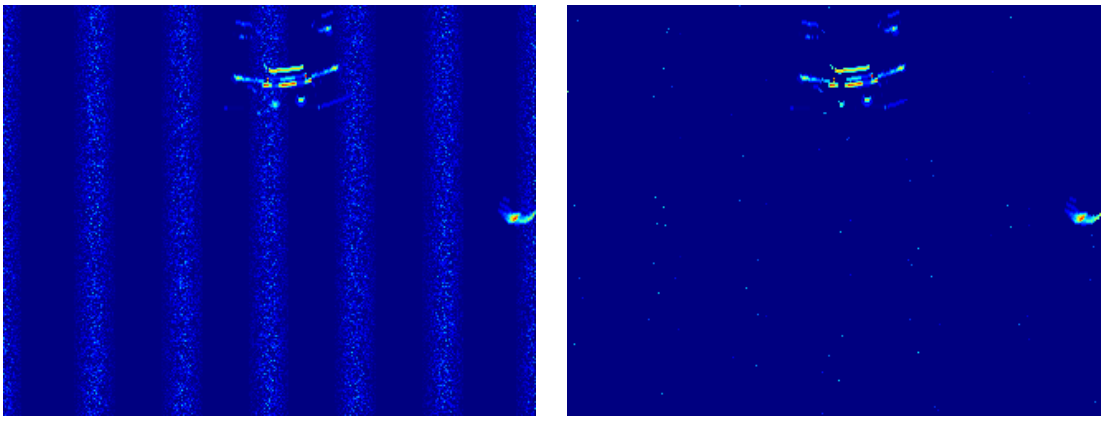
Figures 4.15 and 4.17 showed that this result remained valid when considering sensors of increasingly larger apertures. As a result figure 4.19 exhibited very similar error to coverage values for all sensors at low sampling period (under 5cm). When considering larger sampling periods, an increase in discrepancy in the median error and error to coverage plots is observed, due to the larger blurring effect of large apertures not being compensated by a high sampling resolution along the direction of uncertainty.

It can be observed on figure 4.14 that constant median errors are obtained when using a pencil-beam sensor. While this result might sound counter-intuitive at first, it is due to the fact that the very small aperture of a pencil-beam SONAR (1°) results in a very small vertical 3D spread of the samples in comparison to wider apertures. As a result, when considering large sampling periods, gaps appear between each samples, slowly decreasing the surface coverage but without significantly increasing the final reconstruction error. For this reason, reconstruction of structures featuring a larger proportion of horizontal surface such as the pillar and the bottle exhibit large variations of errors when using a wide-aperture sensor. On the opposite, objects offering no tangential incidence angle such as the sphere lead to very similar reconstruction metrics, independently of the chosen sensor.

Similarly, it can be observed that pencil-beam reconstructions do not feature any outliers due to a low vertical spread. On the contrary, the vertical uncertainty of wide-aperture sensors naturally tend to generate outliers that can only be avoided by sampling at high rate along the direction of uncertainty.

### 4.5.3 Noise corrupted simulation results

In presence of noisy data, a pixel-wise Z-test denoising step is applied before the reconstruction. As illustrated in figure 4.21, applying the Z-test denoising with a given threshold ($T = 3$) enables to suppress most of the noise present in the image by removing a high percentage (99%) of noise points. In opposition to a uniform thresholding of the image, the spatially-varying noise model removes the sensor-specific noise pattern in both high and low SNR regions.



(a) (b)

FIGURE 4.21: Z-test denoising on a simulated image with a threshold T = 3. The Z-test filter removes most of the noise points at the risk of removing data points in areas of low SNR.

As visible in figure 4.22, the quality of 3D reconstructions is affected by the SNR of the input images: the median error remains stable but the surface coverage decreases significantly at low SNR values.

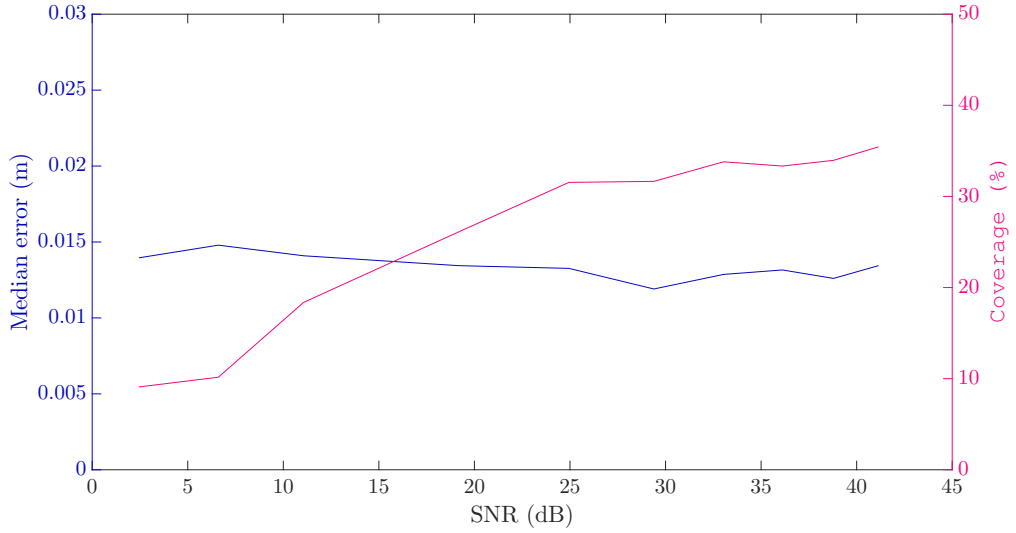



FIGURE 4.22: Median error and coverage versus SNR. Decreasing the SNR of input images does not significantly affect the median error but decreases the surface coverage rate.

The proportion of outliers remains stable when the noise level increases as visible in figure 4.23. As a result of the coverage decrease, the median to coverage ratio follows the input SNR levels in a similar way.

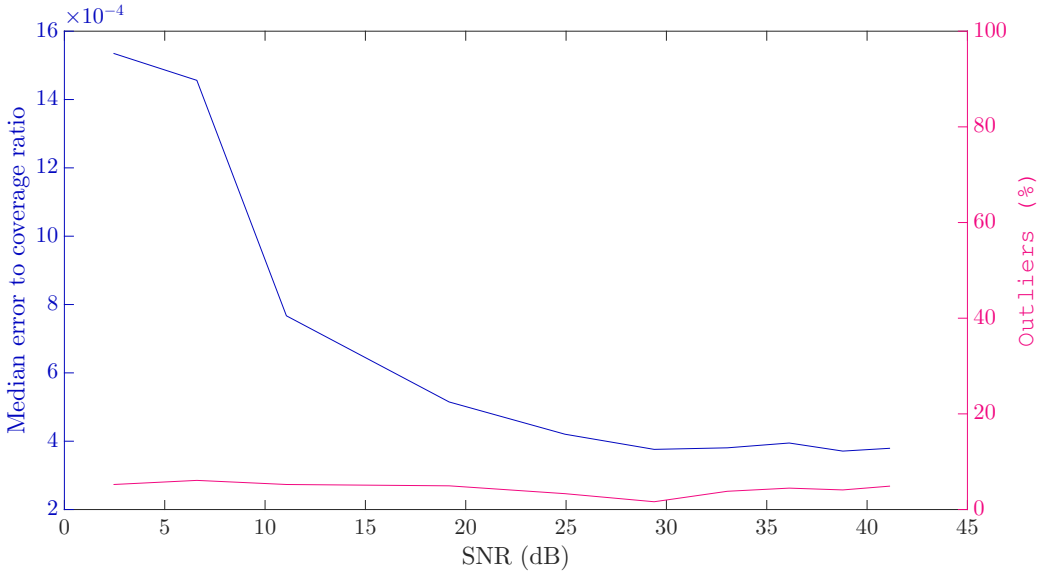



FIGURE 4.23: Median error to coverage ratio and outliers versus SNR. Due to the decrease in coverage at low SNR, the median to coverage ratio follows the same trend as the coverage while the proportion of outliers remains stable (around 5%).

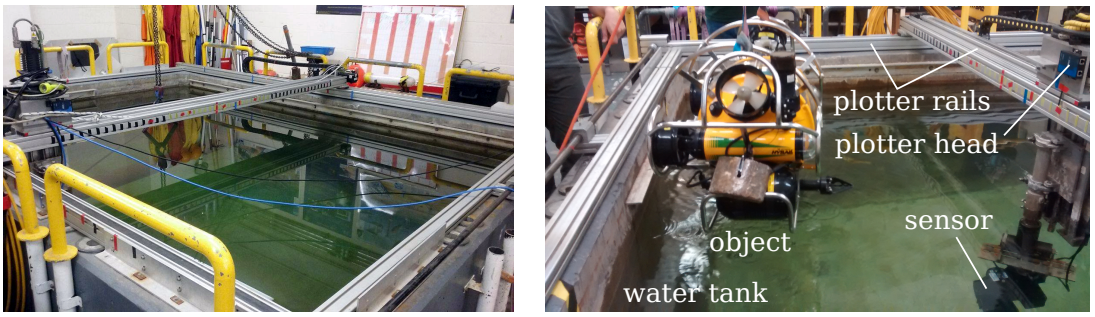
## 4.6 Water tank experimental results

In order to evaluate the performance of the 3D reconstruction algorithm on real SONAR data, we performed tests in Heriot Watt University Ocean Systems Laboratory tank equipped with a sensor positioning system. In comparison to a simulated environment, the SONAR imagery is typically expected to exhibit additional artefacts due to local noise sources and multipaths effects in closed environments. While the imaging quality is therefore expected to be lower, the navigation remains very accurate thanks to the use of an electronic positioning system.

In this section, we first detail here the experimental configuration used for our tests. We then present the 3D reconstruction results obtained with multiple objects.

### 4.6.1 Experimental setup

We gathered real data using two different SONAR heads to scan reference objects that were submerged in a small (4 m × 3 m × 2 m) concrete water tank. As shown in figure 4.24, a 2D plotter mechanism with a programmable interface for motion control, was used to accurately position the sonar head along two axis in the horizontal plane.




(a) (b)

FIGURE 4.24: Heriot-Watt University OSL water tank. a) 4 m × 3 m × 2 m water tank used for experiments. b) 2D plotter mechanism with a BlueView SONAR mounted on its head. The object (Hyball ROV) is being disposed on the bottom of the tank.

The sensors were connected by Ethernet and dedicated ROS nodes were implemented to allow configuration and data reception at a rate of 5 Hz. Similarly, the plotter mechanism provided position readings at a rate of 10 Hz. A step function of constant speed was applied as a command to the plotter, providing uniform samples along the axis of movement. In order to ensure accurate coherency between the position readings and the SONAR images, a synchronization step was applied by applying a bilinear interpolation at the imaging time using the two closest positions (see figure 4.25). We collected data on

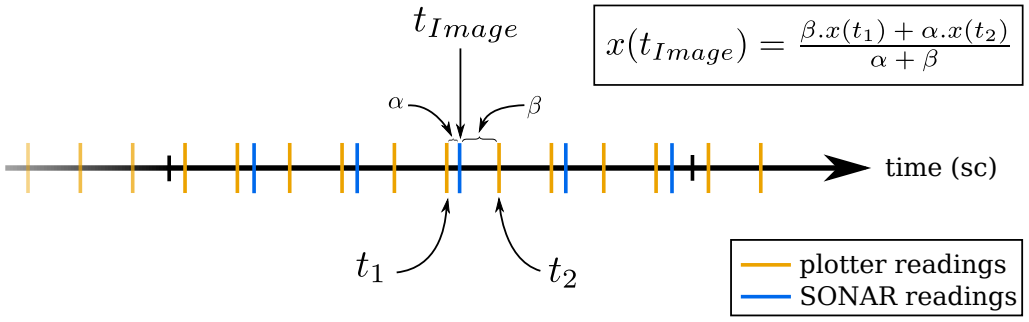



FIGURE 4.25: Synchronization of position and SONAR samples. In order to determine the exact position of the sensor when acquiring the SONAR images, a bilinear interpolation is performed using the two closest positions.

a variety of objects immersed in the tank. For each object, we repeated the experiment with two different sensors mounted on the plotter head: a BlueView MB2250 featuring a small (1°) vertical aperture and an ARIS Explorer 3000 offering a large aperture (14°). As detailed in table 4.2, the two sensors feature very different aperture sizes.

In each case, the sonar was mounted in a downward-looking configuration enabling the acquisition of vertical slices of objects placed on the bottom of the tank. The sampling resolution between successive images was set to 4 mm for both sensors. As illustrated in

| | **BlueView MB2250** | **Aris Explorer 3000** |
|---|---|---|
| **Frequency** | 2.25 MHz | 3.0 MHz |
| **Horizontal aperture ($A_h$)** | 76° | 30° |
| **Vertical aperture ($A_v$)** | 1° | 14° |
| **Beam width** | 1 × 1° | 0.25 × 0.25° |
| **Number of beams** | 256 | 128 |
| **Max. update rate** | 40 Hz | 15 Hz |
| **Power** | 15 Watts | 80 Watts |
| **Depth rating** | 1000 m | 300 m |
| **Size** | 22.6 cm × 11.9 cm × 10.9 cm | 26 cm × 16 cm × 14 cm |
| **Weight in air** | 3.67 kg | 5.17 kg |
| **Weight in water** | -0.09 kg | 1.06 kg |

TABLE 4.2: Specifications of the two sensors used for tank experiments.

table 4.3, the two sensors were configured to feature similar imaging settings: similar bearing and range resolutions are obtained at identical frame rates, the offset in minimum and maximum ranges being due to different sensor sizes and positioning on the plotter head. The plotter shifted the sensor along the long edge of the water tank, over a distance of maximum 2 meters (depending on the size of the object inspected). Acquisitions of complete sequences using both sensors therefore took a maximum of a 100 seconds per object.

| | **BlueView MB2250** | **Aris Explorer 3000** |
|---|---|---|
| **Min / Max FOV** | -38.14° / 38.14° | -15° / 15° |
| **Min range** | 0.25 m | 0.67 m |
| **Max range** | 1.4 m | 1.66 m |
| **Range resolution** | 0.6 cm | 0.3 cm |
| **Bearing resolution** | 0.15° | 0.23° |
| **Image width** | 512 | 128 |
| **Image height** | 174 | 341 |
| **Update rate** | 5 Hz | 5 Hz |

TABLE 4.3: Imaging settings for the two SONAR sensors during the water tank experiments.

As shown in figure 4.26, the objects placed in the tank consisted of an aluminium sphere of radius 15cm, a metal cylinder with 10cm radius and 1.5 m long, a rectangular parallelepiped (1 m × 0.2 × 0.2 m) on top of which a ROV shell was placed irregularly and a Hyball ROV (1 m × 0.5 m × 0.5 m). One or more ropes were attached to each object in order to allow easy deployment and recovery. Although simple, the shapes of the first two objects (figures 4.26-a,b) are frequently found in man-exploited marine environments where buoys and pipelines are commonly used. The two following

(a) (b)

(c) (d)

FIGURE 4.26: Reference objects used for tank experiment. All objects were secured by one of more ropes. a) 30 cm diameter aluminium sphere. b) 1.5 m long × 20 cm diameter. c) 1 m × 0.2 × 0.2 m rectangular parallelepiped with an ROV shell layed on its top. d) Hyball ROV (1 m × 0.5 m × 0.5 m).

configurations (figures 4.26-c,d) exhibit more complex shapes enabling the evaluation of the capacity of the algorithm to reconstruct smaller details. Interestingly, the two configurations feature elements coming from underwater vehicles, therefore simulating a vehicle recovery operation by a SONAR guided system.

### 4.6.2 Experimental results

Reconstruction experiments were carried at 4 mm sampling rate for the BlueView MB2250 and 1.2 cm for the ARIS Explorer 3000. The difference in sampling rate is justified by the difference in aperture. Sub-centimeter reconstructions using the ARIS sensor did not show any significant improvement in the accuracy of the reconstruction. On the contrary, due to the very short range of observation, the 1° aperture Blueview sensor required to be used at low sampling rate to exhibit significant overlap between successive measurements. The results of 3D reconstruction using the carving method are presented in figure 4.27. Figures 4.27-a,c,e,g present the reconstructions obtained from the pencil-beam SONAR (BlueView MB2250) whereas figures 4.27-b,d,f,h show the equivalent reconstructions obtained with a wide-beam SONAR (Aris Explorer 3000).

### 4.6.3 Analysis

While no quantitative analysis is made due to the absence of CAD models, a visual analysis enables to confirm the trends observed in simulations. As could be expected based on the simulation results, the reconstructions from the pencil-beam sensor exhibit better geometrical accuracy with a better ability to reconstruct small details. Similarly the number of outliers is visually larger in wide-beam reconstructions. Although both sensors provided good coverage of the structures due to high sampling rate along the vertical aperture, the reconstruction obtained from the pencil-beam sensor exhibit a few holes. In particular, figures 4.27-a and 4.27-g exhibits more gaps than their equivalent reconstruction from wide-beam samples (figures 4.27-b and 4.27-h).

Although in comparison to simulated results, a lower geometrical accuracy can be observed, the reader should bear in mind the differences in scales between the objects: the sphere used in simulation was featuring a 1m diameter while the pseudo-sphere in the tank is 30cm wide. Similarly the details on the ROV Hyball are typically at centimetre level while the oil-field riser structure presented in simulation featured larger geometrical elements. Although a difference in range resolution could appear to be favourable to the tank configuration with sub-centimetre range resolutions in comparison to 2cm in simulation, these high resolutions are at the upper limit of the resolving capacity of the sensors, coming at the cost of lower resilience to noise.

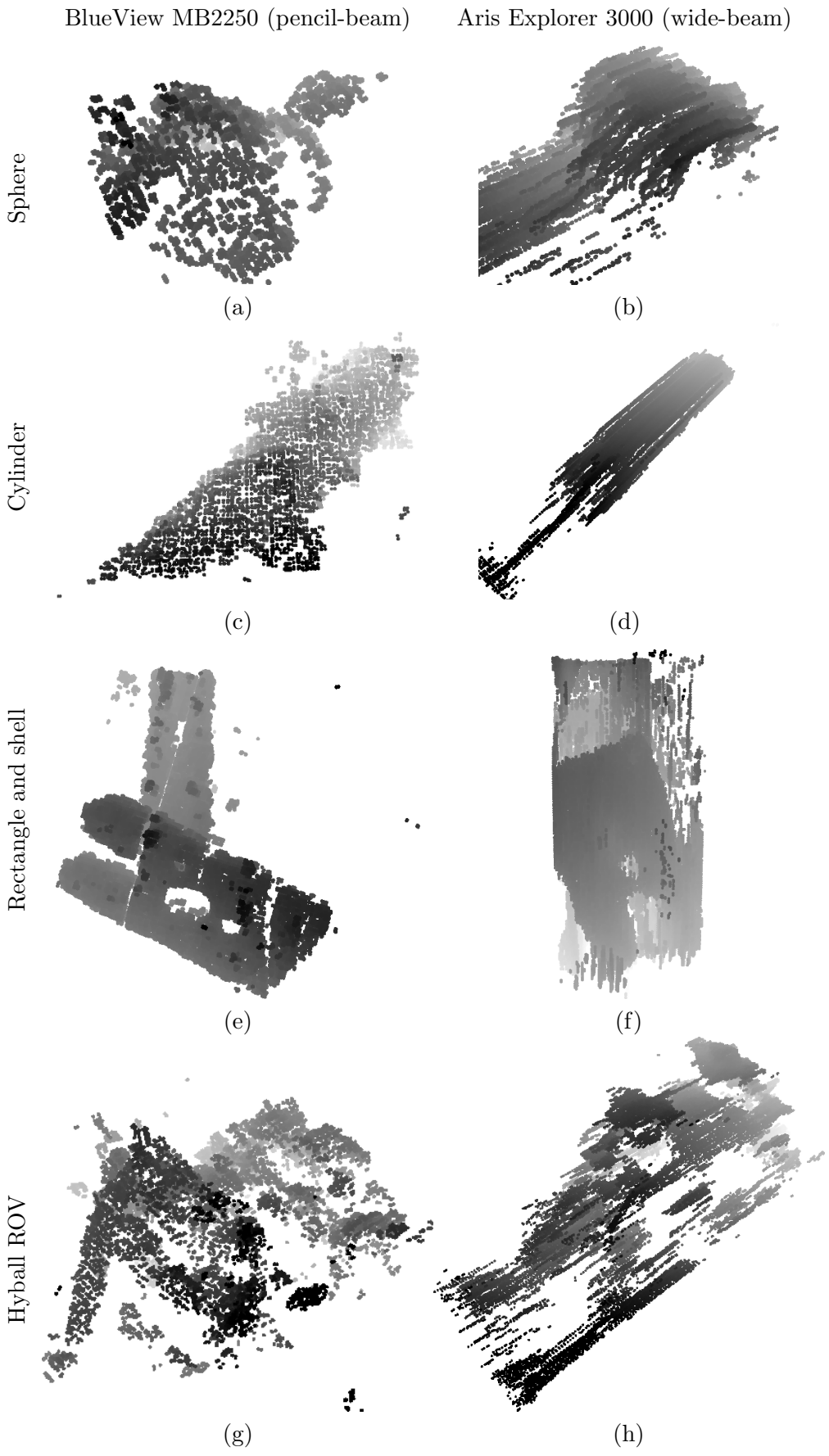



FIGURE 4.27: 3D carving reconstructions from tank data of several reference objects (rows) and two sensors of different vertical aperture (columns).

A further analysis on the differences between the simulated results and real data as well as a comparison between the wide-beam and narrow-beam sensors, are presented in section 5.7.

## 4.7 Field results

In this section, we present 3D reconstruction results obtained using the carving method on two sets of real field data by an Autonomous Underwater Vehicle. In comparison to the data obtained from tank experiment, data acquired on a real underwater vehicle offer the possibility to evaluate the quality of SONAR imaging in an uncontrolled environment and in presence of larger objects. In particular the presence the of marine growth, water pollution and marine wildlife is frequently observed. In addition to this, vehicle positioning and navigation remains a challenge in GPS denied environments such as under water, impacting the estimate of the position of the sensor and therefore the 3D reconstruction.

We first introduce the AUV prototype along with the multiple sensors needed for the data gathering. We then present 3D reconstruction results obtained on multiple oil-field underwater structures.

### 4.7.1 The Autonomous Inspection Vehicle prototype

In partnership with SeeByte, Subsea7 has been developing an AUV adapted to oil field structures inspection and maintenance. This vehicle, called AIV for Autonomous Inspection Vehicle, is designed to achieve up to 24 hours of autonomous surveying and potential interventions at depth of up to 3000 metres. As illustrated in figure 4.28, the vehicle is equipped with two BlueView P900-130 forward looking wide-aperture imaging SONARs (see table 4.4) respectively mounted horizontally and vertically. In addition to this, a pencil-beam BlueView MB2250 SONAR (see specification in table 4.2) is mounted in a downward looking configuration for bathymetry sensing.
Two colour cameras are embedded too, respectively in a forward-looking and a downward-looking configuration. The AIV prototype is equipped with 5 thrusters providing control in $X, Y, Z$ and $Yaw$. The vehicle aims at performing fully autonomous or semi-autonomous mission in both tetherless configuration or via limited acoustic communications. The AIV


FIGURE 4.28: Synchronization of position and SONAR samples. In order to determine the exact position of the sensor when acquiring the SONAR images, a bilinear interpolation is performed using the two closest positions.

| | |
|---|---|
| **Frequency** | 900 kHz |
| **Horizontal aperture ($A_h$)** | 130° |
| **Vertical aperture ($A_v$)** | 20° |
| **Beam width** | 1 × 1° |
| **Number of beams** | 768 |
| **Range resolution** | 2.5 cm |
| **Max. update rate** | 15 Hz |
| **Power** | 19 Watts |
| **Depth rating** | 1000 m |
| **Size** | 31.5 cm × 12.7 cm |
| **Weight in air** | 4.35 kg |
| **Weight in water** | 0.6 kg |

TABLE 4.4: Specifications of the BlueView P900-130 embedded on the AIV.

typically provides structure inspection capability at ranges of 1.5 to 3 metres. The vehicle navigation is obtained from integrating readings from multiple sensors. Bottom-lock velocities are provided by a DVL (Doppler Velocity Log), depth is read from a pressure sensor while orientation estimation is obtained using a compass and a gyroscope. The vehicle is transported in a basket enabling direct deployment on the seabed from a surface vessel. Normal recovery of the vehicle is operated by autonomous docking of the vehicle into the basket. For emergency situations, the vehicle is equipped of satellite communication interface to allow recovery from the sea surface.

The vehicle autonomy framework developed by SeeByte provides mission planning based on a simplified model of the environment called world model. This world model typically features local bathymetry, tidal information and CAD model representations of the

structures of interest. Based on the world model, side or top inspections of the structures can be planned in advanced and simulated with a hardware-in-the-loop simulator. Once the mission started, the vehicle leaves its basket, approaches the structures of interest while avoiding potential obstacles via a SONAR-based collision avoidance module. Once the vehicle located in the vicinity of an underwater structure, online relocalisation is performed using forward-looking SONAR images, enabling side or top structure inspection within safety margins. Within the limits of the navigation system, the structure inspection capability enables both horizontal and vertical pipeline tracking and inspection. Once the mission completed, the vehicle autonomously docks itself into the basket for recovery. Offline data analysis can then be performed allowing video inspection of the structures by an operator.

### 4.7.2 Lake field trials

The first set of trials were conducted in Loch Eil in Fort William, Scotland. On the site of the trials, three metallic structures mimicking typical oil field structures as well as two pipeline sections were placed on the seabed at a depth of around 30 metres. As illustrated in figure 4.29, the structures feature similar dimensions as the structures used in simulation. For clarity, we refer to these structures respectively as the box structure (figure 4.29-a), the brick structure (figure 4.29-b) and the grillage structure (figure 4.29-c).

#### 4.7.2.1 Pencil-beam SONAR inspection

The first experiment consisted in using the downward-looking pencil-beam SONAR located at the bottom of the vehicle to perform a top inspection of each structure present on the field. The along-track sampling period was 4cm with an average distance to the seabed of 5 metres. Following this configuration, the full field was inspected in an hour time.

Using the carving reconstruction technique, each structure has been reconstructed. Figure 4.30 presents the reconstructions of each structure.

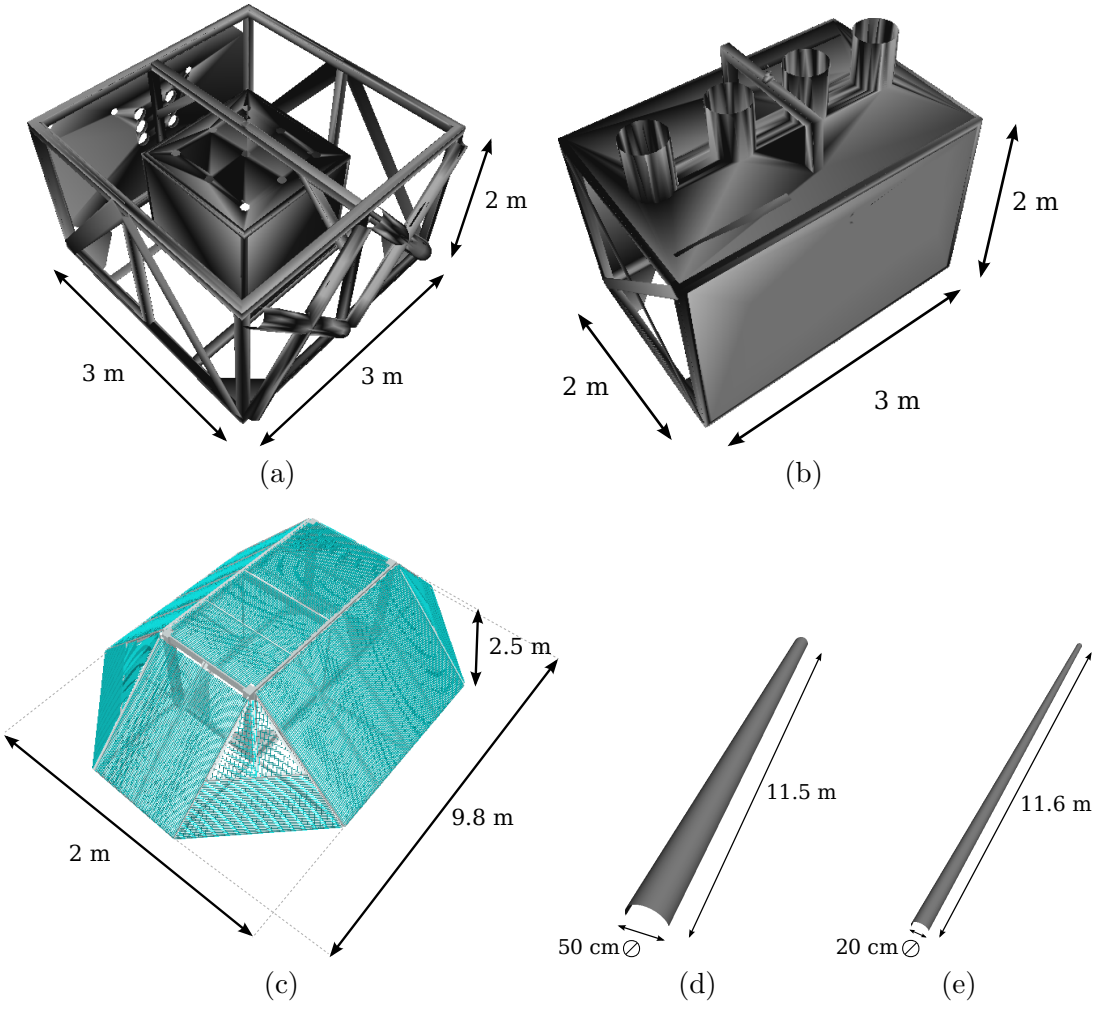



FIGURE 4.29: Lake field trials reference structures. Three metallic structures mimicking oil field structures (the box structure (a), the brick structure (b) and the grillage structure (c)) were deployed on the site of the trials along with two pipelines (d,e).

#### 4.7.2.2 Wide-beam SONAR inspection

In addition to downward-looking pencil-beam images, the data gathered during the AIV trials featured a set of forward-looking views of the box structures with small vertical movement. Figure 4.31 presents a depth map representation of the carving reconstruction of the box structure from a small set of samples (30 images).

### 4.7.3 Offshore trials

A second set of field data was gathered by the AIV during offshore tests including oil field structure inspection, pipeline inspection and seabed mapping. In comparison to in-lake trials, the operation of underwater vehicles is made more difficult offshore due to deeper

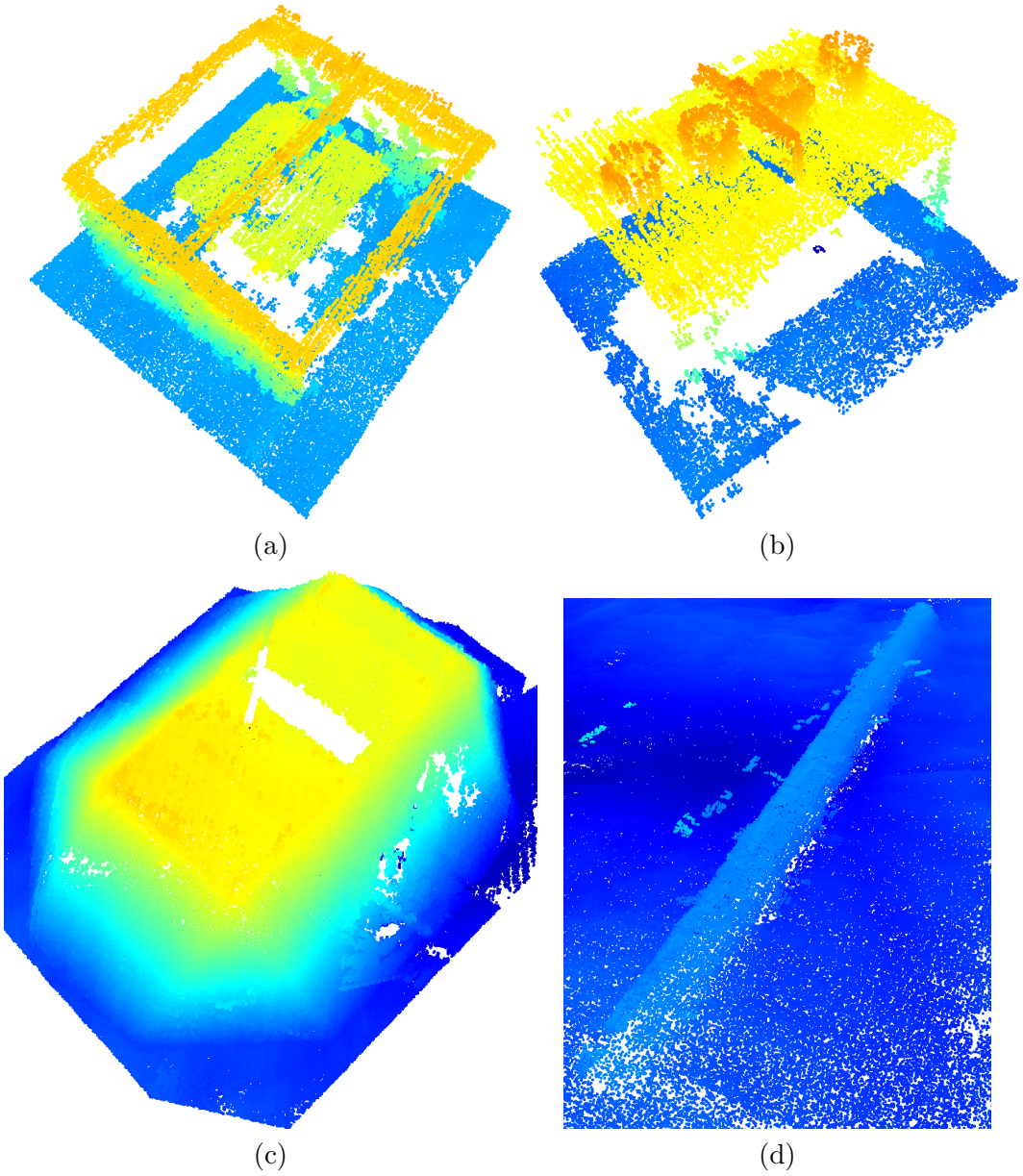
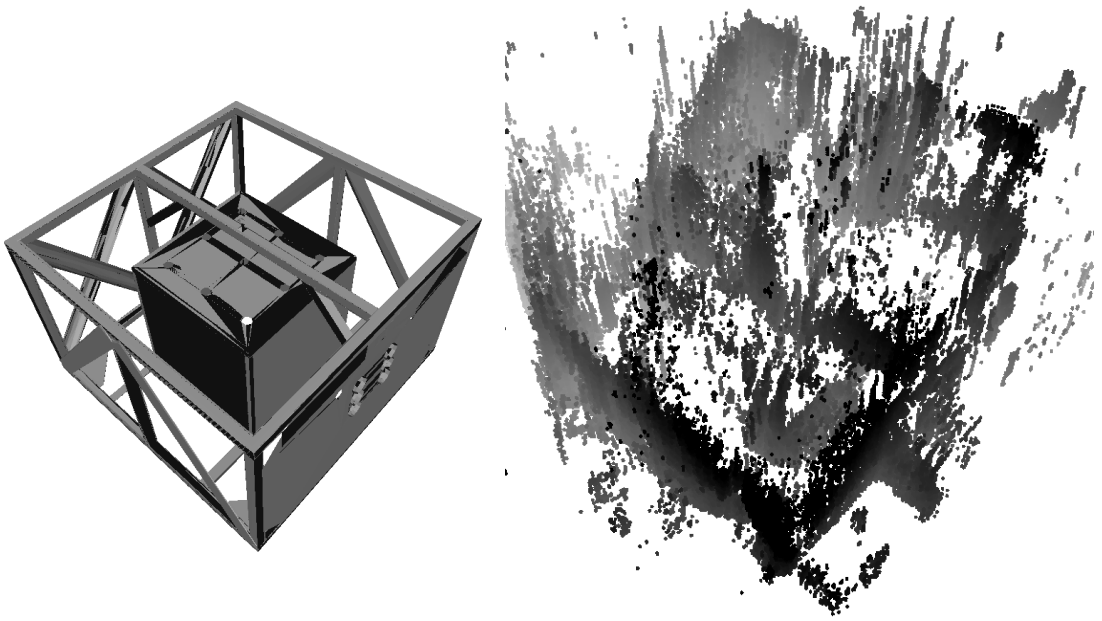



FIGURE 4.30: Loch Eil trials pencil-beam carving reconstructions. 3D carving reconstructions of the box structure (a), the brick structure (b), the grillage (c) and the large pipeline (d). The points are color-coded based on their altitude (Z).

locations, stronger water currents and in our case more advanced marine growth. While the presence of marine growth is not a particular problem for the reconstruction process, it makes detailed assessment of the quality of the reconstruction more complicated as well as any kind of shape-based processing such as automatic tracking and recognition. While the data was gathered using AIV-specific tracking and planning algorithm, the navigation data is accurate enough to perform 3D reconstruction over a few metres.



(a) (b)

FIGURE 4.31: Depth map representation of a 3D carving reconstruction of the box structure from FLS samples (b) in comparison to its CAD model (a).

#### 4.7.3.1 Pencil-beam SONAR inspection

Seabed inspection data using the downward looking was operated above buried pipelines, covered by hard mattresses and sand bags. The inspection was operated at an along track sampling resolution of 3 to 4 cm and at a distance of 1 to 2 metres. Figure 4.32 shows the carving reconstruction of the mattress and the sand bag as well as two video frames of the same objects. Objects of only a few centimetres such as the small rounded rock are noticeable in the SONAR reconstruction and can be visually compared to observations made from a low-resolution underwater camera. The reconstruction exhibits noticeable gaps between each samples due to the narrow vertical aperture of the sensor as predicted in simulation.

#### 4.7.3.2 Wide-beam SONAR inspection

**Polygonal structure**

The data gathered by the AIV during offshore trials also featured forward looking images of a 6 m large and 8m tall polygonal structure with advanced marine growth. The structure was fully inspected during six vertical inspection sections performed at different angles around the structure, covering the full surface of the structure. The average

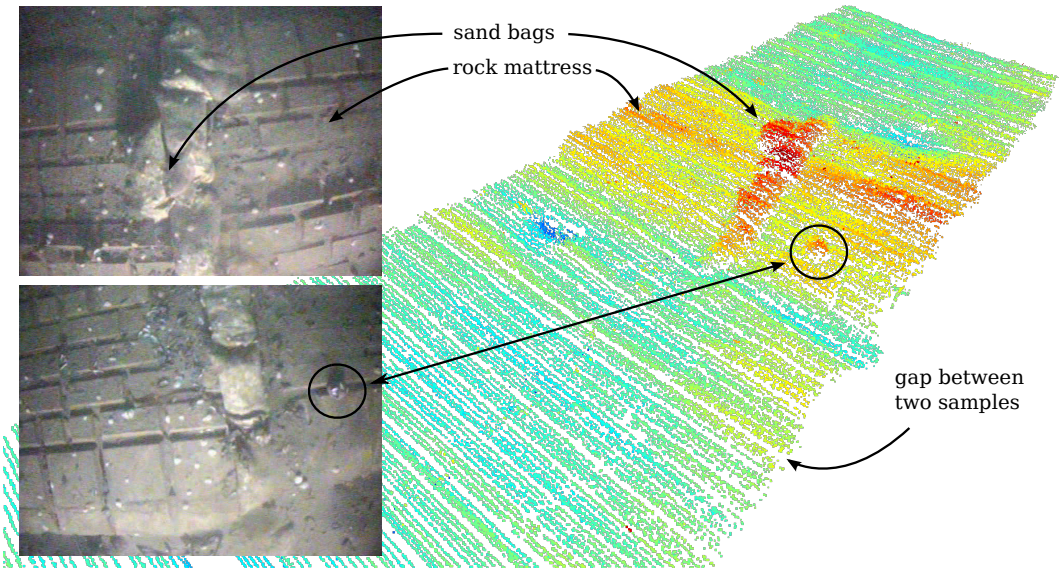



FIGURE 4.32: 3D carving reconstruction of a pipeline and its mattress from pencil-beam samples compared with video data of the same objects.

vertical sampling period was 4 cm and multiple views all around the structure were acquired. The result of a carving reconstruction obtained using the SONAR dataset are shown in figure 4.33. In addition to this, part of one of the faces of the structure appears

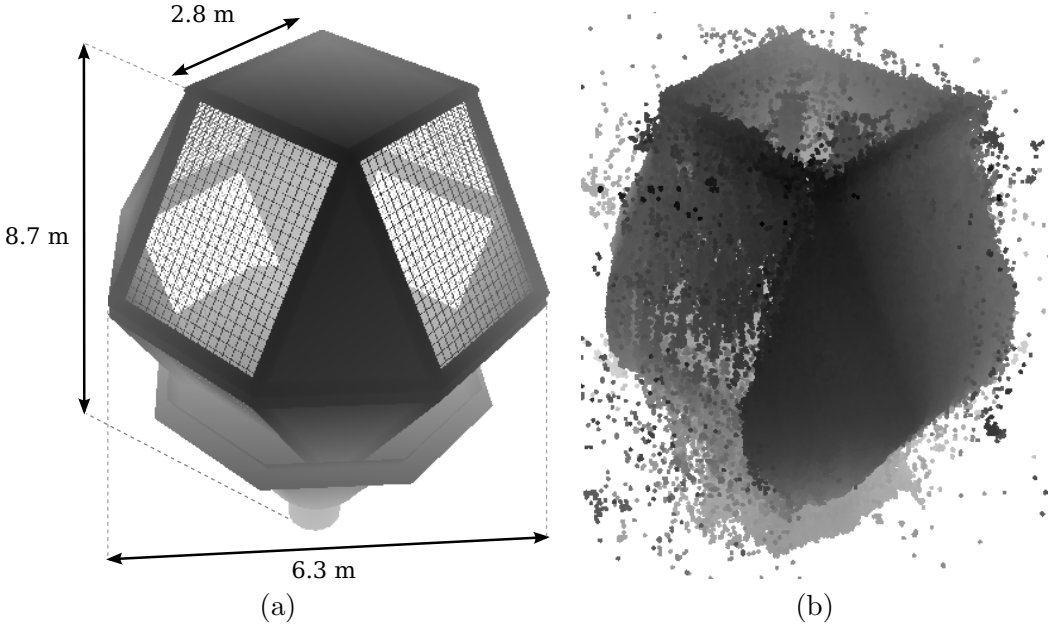



FIGURE 4.33: 3D carving reconstruction of a polygonal structure. a) CAD model of the observed structure, b) reconstructed point cloud using the carving technique.

as missing. The rest of the structure appears densely covered.

### Riser pipeline

Another dataset was gathered while inspecting vertical pipelines called risers. The average vertical sampling period was 7 cm with a distance to the pipeline ranging from 3 to 5 meters. These risers typically feature a 50 cm diameter and are kept buoyant by large buoyancy (1m diameter in this case). In this sequence, the data was recorded over a vertical movement of around 15 m. Figure 4.34 shows a depth map rendering a 3D reconstruction of a riser using the carving technique. The point cloud exhibits two

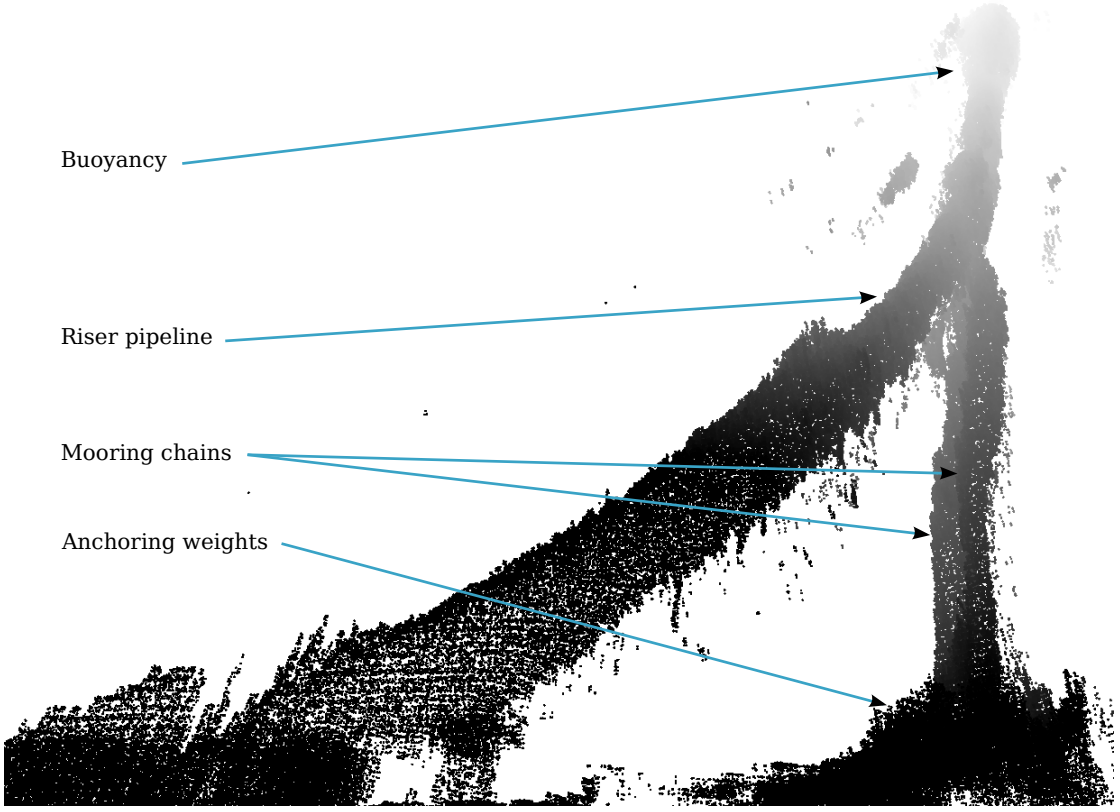



FIGURE 4.34: Depth map of a 3D carving reconstruction of a riser pipeline. The point cloud exhibits the curved riser, two mooring chains going from the riser to the anchoring system. A buoyancy is visible at the top of the point cloud.

mooring chains used for stabilisation of the riser. These chains go from the riser itself to an anchoring point on the seabed. The riser appears as a curved half-cylinder on which buoyancies are placed every couple of meters for hydrodynamical balance, as visible on top of the reconstruction.

### 4.7.4 Computing resource usage

In terms of computational cost, the space carving algorithm processes an image and updates the map every 500ms using an Intel Core i7-4700MQ processor. It is worth noting here that this process is not multi-threaded at the moment, we could therefore expect to be able to process nearly 8 images in this amount of time on this processor. 500ms per iteration remains compatible with a typical underwater structure inspection where vehicles are required to adopt slow motions. After the inspection finished, the occlusion resolution step is run taking an average of one second for each image. This step being multi-threaded (8 threads) it only take 15 seconds to generate the final map from 100 input images. This is an acceptable overall processing time for real operations. Due to the very important number of points being processed and stored (every pixel of an image generates a hundred reprojections along the vertical aperture), up to 16Gb of RAM can be used at the end of a reconstruction with a lot of samples. While this memory usage is relatively high, the reconstruction of small sequences (100 images) takes around 2Gb of memory, therefore remains affordable in an embedded environment. In addition to this, the sequences of input images can always be split up in a few small sets of images, limiting the memory requirements.

### 4.7.5 Analysis

Carving reconstructions on real field experiments exhibited a good level of detail with SONAR sensors of any vertical aperture. The locally-contained navigation drift assumption appears to be valid, enabling accurate reconstruction during an inspection of a few metres. As a result, it is possible to reconstruct large man-made structures such as oil field structures with the resulting reconstructions exhibiting enough detail for human eye recognition of the structure or even identification of small geometrical details such as parts of the structures.

3D reconstructions issued from pencil-beam SONAR data exhibit a lot of details, typically comparable to low resolutions underwater cameras in clear conditions as demonstrated in figure 4.32 but also exhibit decreased coverage in comparison to wide-aperture SONAR data (see figure 4.33). 3D reconstructions with wide-beam SONAR data exhibit a lower level of details and appear to be more sensitive to the sampling scheme as exhibited by

figure 4.31. Provided the sampling along the direction of uncertainty can be performed at sufficiently high rate, the reconstruction results exhibit an increased surface coverage in comparison to pencil-beam SONAR. Importantly, a single wide-beam forward-looking SONAR such as the one embedded on the AIV provided 3D reconstruction of the structure of interest during the inspection of vertical structures while permitting additional mission-specific processing such as 2D tracking of the structures and collision avoidance.

## 4.8 Conclusions

In this chapter, we presented a 3D reconstruction technique from imaging SONAR based on the idea of space carving. First we formulated the reconstruction problem as the estimation of the reflectivity from a set of blurred measurements. Based on this formulation, we derive the concept of space carving which estimates an upper bound to the reflectivity by considering the minimum intensity observed at each point, scaled by the radiating pattern of the sensor.

We then presented a three steps space carving algorithm enabling the 3D reconstruction to be performed during the acquisition of new samples. This carving algorithm enables to address the occlusion phenomenon occurring during the SONAR imaging process by performing a so-called conservative spherical reprojection and performing a last pass to explicitly associate each observation to sections of the vertical aperture. In addition to this, our algorithm achieves efficient storage by operating a 3D scaling between the irregular sampling inherent to SONAR imaging model and the regular sampling grid of an Octree structure.

In order to evaluate the performance of the reconstruction algorithm, tests in a simulation environment were performed on a set of reference objects mimicking man-made objects frequently found in a marine environment. A quantitative analysis of the reconstruction results was first performed on noise-free data with two sensors of different aperture, respectively representing a pencil-beam SONAR and a wide-beam SONAR. Importantly, four metrics were used to provide a complete view of the reconstruction quality by not only considering the geometrical error but also evaluating the surface coverage of the inspected object as well as the presence of outliers in the final point cloud. The quantitative analysis showed that the uncertainty resulting from the vertical aperture of the sensor could be addressed by sampling at high enough rate along the direction of uncertainty. In particular, when considering small enough sampling periods, similar reconstruction quality could be obtained with both a pencil-beam and a wide-aperture SONAR: while pencil-beam SONARs naturally provide lower uncertainty on the Z axis, the small increase in error occurring when using wide-aperture SONAR is balanced by an increased surface coverage of the structure.

We then evaluated the carving reconstruction technique on real data. First, data acquired in a water tank with four small reference objects and two different SONAR heads was used to evaluate the quality of reconstruction in a challenging, closed environment, leading to SONAR images featuring increased levels of noise and multi-path effects. The four objects were successfully reconstructed and appear as recognisable using both the pencil-beam and the wide-beam sensor. In order to test the reconstruction method in a real operation environment, field data was gathered by an AUV prototype in two different locations. While the first set presented in-lake field data featuring low water currents and marine growth-free structures, the second dataset was gathered on a real offshore oil field providing inspection data of structures such as pipelines, risers and typical oil field structures. Water tank experiments exhibited the difficulty of reconstructing small objects with a high level of details when operating in acoustically challenging environments whereas real field data tests showed good reconstruction accuracy with enough details to allow for clear human-eye recognition of the structures. Similarly, tests made with both a pencil-beam and a wide-beam imaging SONAR showed the improvement in resolution when using pencil-beam SONAR but also showed the viability of using a wide-beam SONAR to recover 3D shapes of large objects with a slightly lower level of details. In particular, our field results showed the importance of the along-track sampling rate and the advantage of wide-aperture SONARs when operating at low sampling rates.

# Chapter 5

# Reconstruction as a deconvolution

## 5.1 Introduction

In chapter 4, we presented a non-linear reconstruction method based on a space carving approach. This technique allowed an iterative, per-sample reconstruction of the scene from observations taken at different positions along the direction of uncertainty without strict constraints on the sampling positions. On the contrary and as detailed in chapter 3, SAS methods (Sæbø et al. [2013]) traditionally rely on very strict sampling scheme with rectilinear and uniform motion (to a fraction of wavelength accuracy) making their use on AUVs often impractical. Although 2D SONAR do not provide direct phase information, therefore prohibiting the use of SAS techniques, the reconstruction from multiple overlapping observations is common to our aperture problem. Model-based methods such as Coiras et al. [2007] showed their interest in estimating multiple parameters from a model-to-data iterative matching.

Inspired by this, we propose in this chapter to model the observation process using a rectilinear and uniform sampling scheme to reformulate the 3D reconstruction problem as a deconvolution.

Based on the SONAR imaging model, we first reformulate the observation process along a vertical line as a convolution with an unknown and spatially-varying kernel and exhibit the illness of the reconstruction problem. Through a short review, we provide references on the state of the art of deconvolution techniques. In order to address the complexity of the deconvolution process, simple assumptions are made, enabling the formulation of

the reconstruction problem as a constrained sparse linear system optimization. Similarly to the space carving algorithm, experiments made on simulated data as well as two real datasets are presented and analysed. We demonstrate the interest and limitations of adding regularization to the deconvolution method. We finally present a complete analysis of the results with a particular focus on comparing the space carving and the deconvolution methods, as well as a comparison between the use of pencil-beam and wide-beam sensors for 3D reconstructions.

## 5.2 Formulation as a spatially-variant blind deconvolution problem

In this section, we formulate the imaging process as a spatially-variant convolution. Inspired by synthetic aperture imaging, we propose to obtain the images by relocating the source along a direction orthogonal to the range as well as the sensor array. Without loss of generality, we depict this chosen direction as aligned with the vertical Z-axis, as illustrated in figure 5.1-a. When acquiring multiple samples along a direction orthogonal

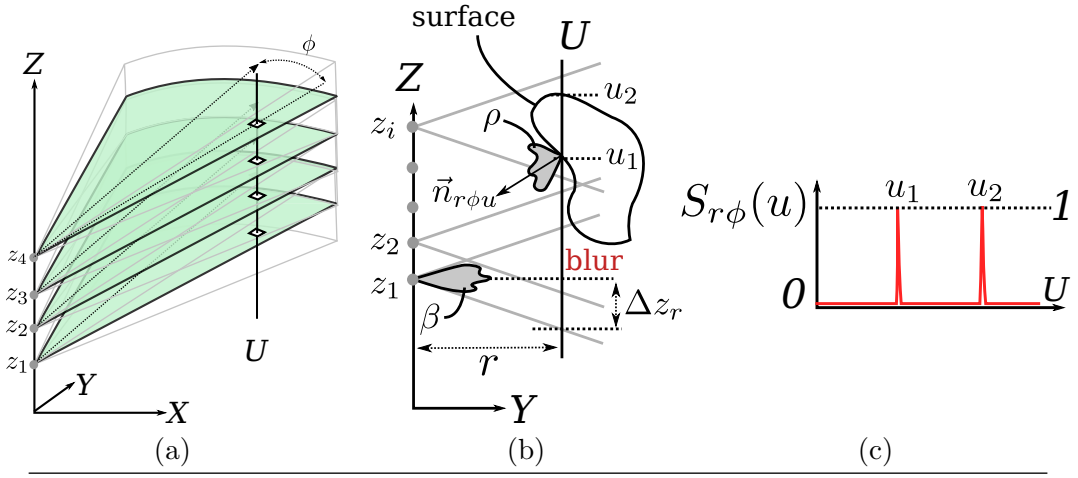



FIGURE 5.1: SONAR imaging model as a convolution along the U axis. When acquiring multiple samples along a direction orthogonal to the image plane (a), the measuring process at range $r$ and bearing angle $\phi$ can be seen as a spatially variant convolution along the U axis (b) with a beam pattern attenuation ($\beta_r$), the vertical aperture of the sensor blurring the observations by summation (integration) along the U axis and a reflectivity $\rho$ depending on the local normal $\vec{n}_{r\phi u}$ and the point of view, the observations being then integrated on the surface indicator function $S_{r\phi}$ (c).

to the image plane, the measuring process at range $r$ and bearing angle $\phi$ can be seen as a spatially-variant convolution along the U axis (see figure 5.1-b) with a beam pattern attenuation $\beta_r$ scaling the returns along the vertical aperture of the sensor. The vertical aperture blurs the observations by integration along the U axis. The reflectivity $\rho$ depends on the local normal $\vec{n}_{r\phi u}$ as well as the point of view making the convolution spatially variant. Finally, as depicted in figure 5.1-c, the observations are then integrated on the surface indicator function $S_{r\phi}$.

By tailoring equation 2.13, we reformulate the imaging model as a convolution:

$$I_{Sonar}(r,\phi) = I_0 \int_{A_v} B_v(\theta)^2 \, S_s(r,\theta,\phi) \, \mu(r,\theta,\phi) \, \frac{\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}}{\|\vec{v}_{\theta\phi}\| \, \|\vec{n}_{r\theta\phi}\|} \, \mathrm{d}\theta$$

First, we introduce the $Z$-coordinate $z_i$ of the sensor array. We also assume that the SONAR wavefronts arriving at the scanned geometry are close to planar (far-field assumption - Aykin and Negahdaripour [2013]). Under this assumption, the integration domain is approximated to be parallel to the $Z$-axis. Finally, we replace the cosine term $(\frac{\vec{v}_{\theta\phi}.\vec{n}_{r\theta\phi}}{\|\vec{v}_{\theta\phi}\| \, \|\vec{n}_{r\theta\phi}\|})$ with a general reflectivity function $\rho$ which subsumes the cosine. Thus, equation 2.13 transforms into equation 5.1:

$$I(z, r, \phi) = I_0 \int_{z-\Delta z_r}^{z+\Delta z_r} \beta_r(u - z) \; S(u, r, \phi) \; \rho(z, u, \vec{n}_{r\phi u}) \;\; \mathrm{d}u \tag{5.1}$$

where $\Delta z_r \approx r \tan \frac{\theta_2 - \theta_1}{2}$ with $A_v = [\theta_1, \theta_2]$. $\beta_r$ is the warped beam pattern over the wavefront at distance $r$, $\rho$ is the reflectivity function evaluated for retro-reflection from the source to the point at $u$ where the surface has a normal $\vec{n}_{r\phi u}$ and $S(u, r, \phi)$ is a binary function that is unity if there is a surface at a distance of $r$ from the source, along the bearing $\phi$ and offset by $u$ in the Z-direction and zero otherwise. Although the coordinate system appears confusing due to a mixture of cylindrical and Cartesian coordinates, recall that $r$ and $\phi$ are common to all images taken by relocating the source along the $Z$-axis. $z$ represents the height at which the source is placed and $u$ is the height at which the reflector is located.

The variation in the measured images at some fixed pixel, $I_{r,\phi}$ as a function of the source position $z$ is:

$$I_{r\phi}(z) = \int_{z-\Delta z_r}^{z+\Delta z_r} \beta_r(u - z) \; S_{r\phi}(u) \; \rho(z, u, \vec{n}_{r\phi u}) \;\; \mathrm{d}u \tag{5.2}$$

where $S_{r\phi}(u)$ corresponds to an indicator function that is unity if there is a surface at height $u$ corresponding to the range $r$ from the Z-axis and bearing $\phi$ wrt $\vec{v}$ (see the plot in figure 5.1). For a given sensor, $\beta_r$ is known but the following two terms in the integral depend on the geometry (surface indicator function and surface normal respectively).

Equation 5.2 suggests that, for a set of SONAR images acquired along the $Z$ direction, the values at each pixel over $z$ represent a convolved version of the surface indicator function. The kernel of the convolution has two components: a fixed component ($\beta_r$) which is known (bessel function described in section 2.2.8) and $\rho$ a spatially-varying and data-dependent modulating term defined by the BRDF at the angle of observation (incidence angle relatively to the surface normal). Since we assume the reflectivity

distribution as diffuse, $\rho$ follows the lambertian model (cosine term described in section 2.2.4) but its estimation remains spatially-variant (over $z$). While the direction of arrival of the emitted sound pulse is known, the shape and position of scatterers in the scenes is a-priori unkown making the surface normal at the location of each acoustic unkown. Therefore without any knowledge on the shape of the object and due to the specifity of the SONAR imaging process (dependency on the normals) the convolution kernel is unknown. Reconstructing surface points S from the sole observations of $I_{r\phi}(z)$ is *therefore equivalent to a blind-deconvolution where the kernel is spatially-varying.*

## 5.3 Review on deconvolution methods

In this section, we present an overview of state-of-the-art deconvolution methods as well as their limitations. Similarly to the SONAR imaging process, we consider the convolution of an input signal by a kernel $h$. This kernel is commonly represented by a point spread function which is the result of a convolution of this kernel with an impulse. The point spread function is so-called for its typical blurring behaviour where an impulse is converted in a wider (lower frequency) output signal, effectively spreading the energy of the incoming signal. Similarly, the convolution process described in section 5.2 blurs the acoustic returns coming from different elevations angles through an integration process. For this reason, the convolution kernel is referred here as a blurring kernel.

### 5.3.1 Fourier-based inversion

When considering the following convolution of an input function $f$ by a blurring kernel $h$ resulting in observations $g$:

$$g(t) = (h * f)(t) = \int_{-\infty}^{+\infty} h(t-u)\ f(u)\ du \tag{5.3}$$

and its associated formulation in the frequency domain:

$$G(w) = \mathcal{F}[g(t)] = \mathcal{F}[(h * f)(t)] = H(w).F(w) \tag{5.4}$$

A natural solution for the estimation of $f$ is then given by a direct Fourier inversion:

$$\hat{f}(t) = \mathcal{F}^{-1}[F(w)] = \mathcal{F}^{-1}\left[\frac{G(w)}{H(w)}\right] \tag{5.5}$$

While simple, this solution generally leads to visible artefacts and noise amplification due to the lowpass behaviour of the blurring function $h$, resulting in low Fourier coefficients at high frequencies $w$.

When an estimate of the noise can be obtained, the following model is assumed:

$$g(t) = (h * f)(t) + n(t) \tag{5.6}$$

The Wiener filter provides an optimal deconvolution filter by minimising the least square error between $f$ and $\hat{f}$, leading to:

$$\hat{F}(w) = D(w).G(w) \tag{5.7}$$

with

$$D(w) = \frac{H^*(w)}{|H(w)|^2 + K(w)} \tag{5.8}$$

and

$$K(w) = \frac{|N(w)|^2}{|F(w)|^2} \tag{5.9}$$

While the Wiener filter provides the optimal trade-off between noise attenuation (smoothing) and inverse filtering, it requires an estimate of both the blurring function $h$ and the noise characteristics $n$.

### 5.3.2 Bayesian inference and MAP formulation

In practice, the blurring kernel $h$ and the noise model $n$ are often modelled mathematically making particular approximations based on the application. In order to take into account the approximations made in the modelling, the reconstruction of $f$ can be formulated by the posterior distribution of $f$ given the observations of $g$. Using Bayesian inference, this distribution can be expressed as:

$$p(f \mid g) \;=\; p(g \mid f)\,\frac{p(f)}{p(g)} \;\propto\; p(g \mid f)\,p(f) \tag{5.10}$$

where $p(g \mid f)$ represents the likelihood of observing $g$ with $f$ as an input and $p(f)$ represents a prior assumption on the distribution of $f$. This simple formulation offers the possibility to obtain the distribution of the function to be reconstructed by modelling the convolution process in the likelihood term and the estimated characteristics of the signal to be reconstructed in the prior information term.

#### 5.3.2.1 MAP formulation

Based on this formulation an estimate of the optimal $f$ is given by a Maximum A Posteriori (MAP) estimation:

$$\hat{f}_{MAP} \;=\; \arg\max_{f} p(f \mid g) \;=\; \arg\max_{f} p(g \mid f)\, p(f) \tag{5.11}$$

In practice, $p(g \mid f)$ and $p(f)$ are commonly assumed to be convex functions $\psi_{Likelihood}$ and $\psi_{Prior}$ such that:

$$\psi_{Likelihood}(f) \;=\; \exp(-\|g - Af\|_p) \tag{5.12}$$

representing the difference between the observations $g$ and the blurring operator $A$ modulating the input $f$ and:

$$\psi_{Prior}(f) \;=\; \lambda\, \exp(-\|f\|_k) \tag{5.13}$$

with $p$ and $k$ defining the norms to model the probability distributions.

In this context, estimating the maximum value of the probability is equivalent to estimating the minimum of its negative logarithm (log-likelihood):

$$\hat{f}_{MAP} \;=\; \arg\max_{f} \psi_{Likelihood}(f).\psi_{Prior}(f) \;=\; \arg\min_{f} \|g - Af\|_p + \lambda\, \|f\|_k \tag{5.14}$$

Adopting a MAP formulation, the deconvolution then amounts to minimizing a linear system modelling the convolution process and making assumptions on the function $f$ to be estimated. Due to its capacity to model the properties of the solution, independently of the likelihood convergence, the prior term is often called regularization term while the likelihood term is frequently referred to as data fidelity term. The choice of $p$ and $k$

then determines the smoothness or the sparsity of the estimated optimum. Due to its interesting properties (differentiability and smoothness promotion), a $L_2$ norm is often chosen for the data fidelity term resulting in the following MAP formulation:

$$\hat{f} = \arg\min_f (g - Af)^2 + \lambda \|f\|_k \tag{5.15}$$

where the estimated function $\hat{f}$ is obtained by least square minimization akin to an energy minimization. The choice of $k$ then leads to various regularization schemes.

#### 5.3.2.2 Regularization

A common regularization strategy is to employ a $L_2$ norm loss term, effectively selecting the solution of minimum energy:

$$\hat{f} = \arg\min_f (g - Af)^2 + \lambda \|f\|_2 \tag{5.16}$$

This regularization scheme, often referred to as Tikhonov regularization (Tikhonov et al. [2013]) or ridge regression (Hoerl and Kennard [1970]) offers the interest of having an analytical solution, therefore not requiring to apply any optimization algorithm.

While the Tikhonov regularization term typically imposes smoothness on the solution, sparsity is frequently desired in imaging problems where sharp transitions (edges) in the original signal are assumed. In this context, a natural sparsity operator is the $L_0$ norm, equivalent to the number of non-zero elements in f:

$$\hat{f} = \arg\min_f (g - Af)^2 + \lambda \|f\|_0 \tag{5.17}$$

It can be showed that this minimization problem is a NP-hard problem (Natarajan [1995]) making its resolution costly if not impossible.

In order to address this problem, $L1$ minimization also called basis pursuit or Lasso minimization (Tibshirani [1996]) is often used to impose sparsity:

$$\hat{f} = \arg\min_f (g - Af)^2 + \lambda \|f\|_1 \tag{5.18}$$

Although convex, the $L_1$ norm remains non differentiable at 0 but the use of a proximal operator enables to reformulate the problem under a smooth (therefore differentiable) approximation around the corner point.

Other regularization operations based on the gradient of the function have been studied. In particular, the total variation regularization (Rudin et al. [1992]) aims at imposing sparsity on the gradient of the function:

$$\hat{f} = \arg\min_{f} (g - Af)^2 + \lambda \left\| \nabla f \right\|_1 \tag{5.19}$$

therefore imposing sparsity on the number of transitions in the input signal. As a consequence, the use of total variation regularization results in a piecewise constant solution often adapted to describing quantized signals such as images.

In any regularization strategy, the regularization term is weighted by the scalar $\lambda$ dedicated to balance the data fidelity term $g - Af$ and the regularity of the solution. While setting $\lambda$ to zero is equivalent to consider the problem as perfectly modelled, regularization-free problem, in most cases the $\lambda$ coefficient needs to be determined empirically. As a consequence, while regularization is often desirable in presence of noise corrupted observations or approximate modelling of the problem, its practical use requires some tuning. Therefore and considering the 3D reconstruction problem from potentially different imaging sources (different sensors), the need for regularization needs to be traded with the necessity of a fully automated processing reconstruction technique.

### 5.3.3 Blind deconvolution methods

When the blurring kernel is unknown, the deconvolution problem is referred to as a blind deconvolution. In this situation, both the original signal $f$ and the blurring matrix $A$ need to be estimated:

$$\hat{f} = \arg\min_{f,A} (g - Af)^2 + \lambda \left\| f \right\|_k \tag{5.20}$$

This formulation results in an inherently ill-posed problem. When the blurring kernel cannot be estimated directly from the observations or external measurements (Joshi et al. [2010]), iterative methods can be employed where A and f are iteratively refined one after the other. Although widely investigated, a recent review on blind deconvolution methods (Levin et al. [2009]) highlighted the limitations of current methods as favouring no-blur

kernels and recall that most of the current methods assume spatially-invariant blurring (Cannon [1976], Chan and Wong [1998], Krishnan et al. [2011]) which often appears to be unrealistic. As discussed in Harmeling et al. [2010], spatially-variant blind deconvolution is a much more complex problem than spatially invariant deconvolution, to which no generic solution has been obtained so far. More generally, blind deconvolution problems are addressed in an application-specific way by making assumptions on the shape of the kernel.

## 5.4 Sparse linear system of the 3D reconstruction problem

We address the 3D reconstruction deconvolution problem using a sparse linear system formulation. For $1 \leq i \leq N$, let $z_i$ be the $Z$ coordinate of the sensor location for the $i^{th}$ acquisition and let $I(z_i)$ denote the corresponding image. Figure 5.2 illustrates the geometry of a measurement at a range $r$ and a range resolution $\Delta_r$. We observe that the

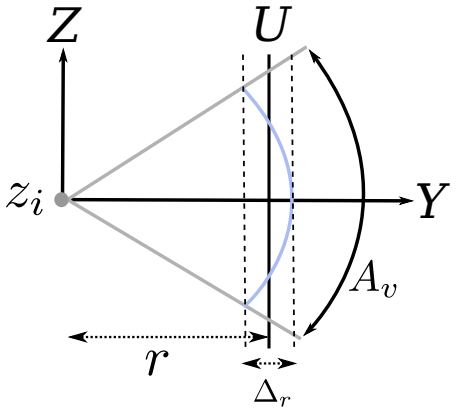



FIGURE 5.2: Validity of the rectilinear approximation based on the vertical aperture $A_v$ and the range resolution $\Delta_r$.

points lying along the vertical aperture can considered as aligned on a vertical axis when the following condition is satisfied:

$$A_v \leq 2 \arccos\left(1 - \frac{\Delta_r}{2r}\right) \tag{5.21}$$

In further sections, this approximation will be refereed to as rectilinear approximation and will be studied in section 5.6. Typical observation ranges considered in this study are around 2 metres and a standard range resolution for imaging SONAR is 2cm. Based on equation 5.21, the rectilinear approximation is valid for apertures smaller than 22°

which in practice coincides with the applications presented in this study. As a result we consider that the vector of values $\{I_{r\phi}(z_i)\}_{1\leq i\leq N}$ corresponds to measured reflectivities of the surfaces along a line parallel to the $Z$-axis at a distance of $r$ from it. We call this the $u$-axis, since it corresponds to the variable of integration $u$ in equation 5.2.

Let the $u$-axis be discretized into $P$ segments. $P$ may be used as a parameter that controls the approximation. For large enough $P$, we assume that the variation of all three components of the integrand in equation 5.2 are negligible within each of these segments and that they are mutually uncorrelated. Then, the integral can be replaced by a summation:

$$I_{r\phi}(z_i) \quad = \quad \sum_{j=1}^{P} \tilde{\beta}_j^i \; \tilde{\rho}_{jr\phi}^i \; \tilde{S}_{jr\phi} \tag{5.22}$$

where $\tilde{\beta}_j^i$, $\tilde{\rho}_{jr\phi}^i$ and $\tilde{S}_{jr\phi}$ are expected values of $\tilde{\beta}_r(u)$, $S_{r\phi}(u)$ and $\rho(z, u, \vec{n}_{r\phi u})$ over the $j^{th}$ segment on the $u$-axis relative to the sensor placed at $z_i$. Given $N$ images, we can then write:

$$I \quad = \quad A_{N\times P} \,.\, b \tag{5.23}$$

where $A$ is a matrix with $A_{ij} = \tilde{\beta}_j^i \tilde{\rho}_{jr\phi}^i$ , $b$ is a vector with $b_j = \tilde{S}_{jr\phi}$, $j = 1, 2, 3, .., P$, which indicates the presence of a surface in the $j^{th}$ segment along the $u$-axis. Since the beam pattern is only non-zero for an interval corresponding to the width of the aperture projected onto the $u$-axis, the matrix $A$ is sparse and the vertical footprint of each measurement on the U-axis only contains $K$ segments.

A graphical interpretation of the linear system is presented in figure 5.3 where 6 measurements are taken along the Z-axis to estimate the surface indicator values on 8 segments along the U-axis. In this situation, each $i$-th measurement at range $r$ can be subdivided in 3 segments and represented as the sum of the $b_j$ coefficients with $i \leq j \leq i + K - 1$ weighted by their associated beam pattern coefficients $\tilde{\beta}_j^i$ and reflectivity coefficients $\tilde{\rho}_{jr\phi}^i$.

While $P$ defines the resolution of the reconstruction along the U-axis, in practice its choice is driven by the number of observations available at a given range (see figure 5.4). In a similar way to Shannon sampling theorem (Shannon [1949]) and in absence of additional information on the function to be reconstructed, the maximum resolution

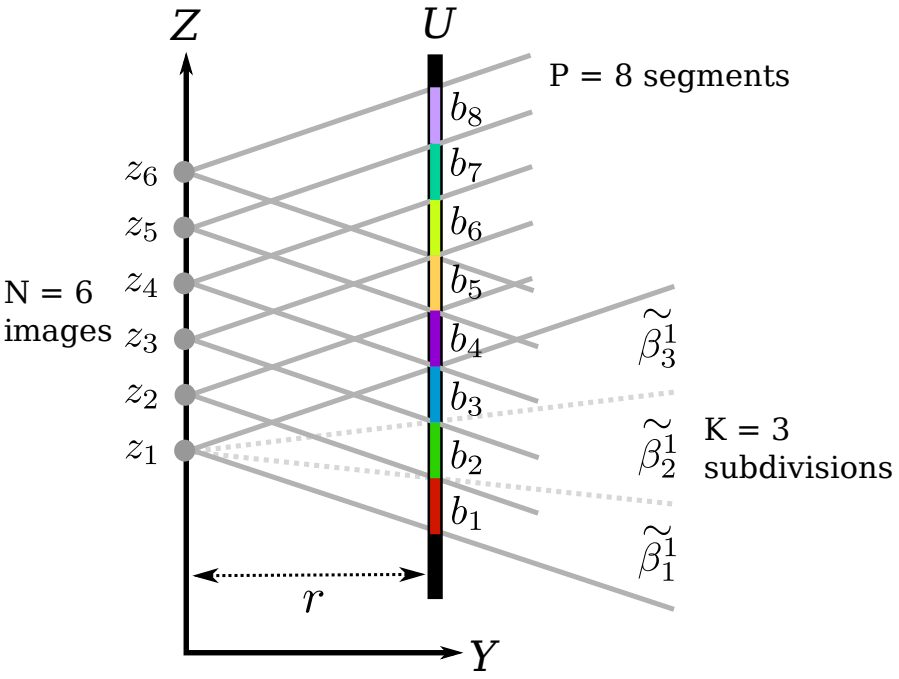



FIGURE 5.3: Graphical representation of the linear system formulation. The surface presence coefficients of $P = 8$ segments are estimated by describing the measurement process of $N = 6$ images through a linear system.

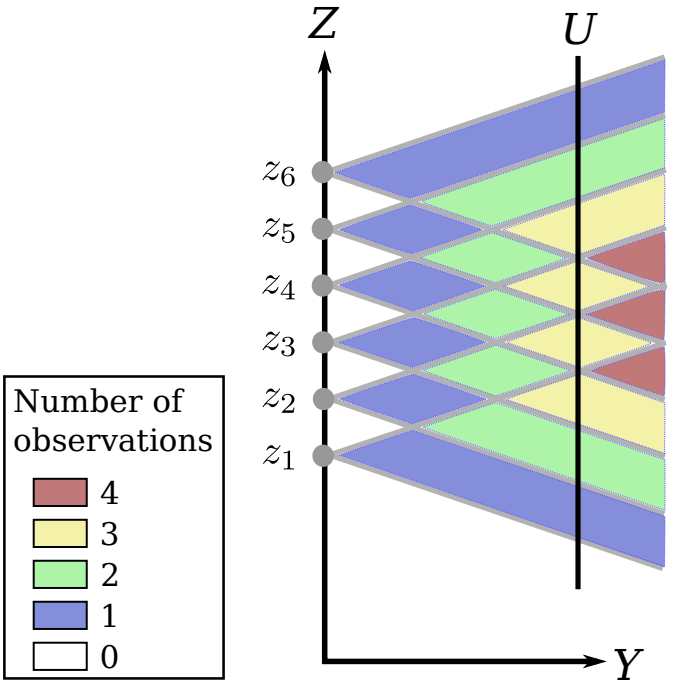



FIGURE 5.4: Range dependent vertical resolution. For a given vertical aperture and sampling period along the vertical axis, the number of observations at a point is range-dependent.

along the vertical axis is limited by the number of observations available for the point considered. Given a vertical sampling period $\Delta_z$ and a vertical aperture $A_v$, it can be shown graphically that the minimum range $r_{min}$ for which $K$ observations are available satisfies:

$$r_{min} = \frac{K\ \Delta_z}{2\ \tan(\frac{A_v}{2})} \tag{5.24}$$

As a result and in order to reconstruct at a range $r$, we choose $K$ as:

$$K = \left\lfloor \frac{2\ r\ \tan(\frac{A_v}{2})}{\Delta_z} \right\rfloor \tag{5.25}$$

In this situation, the number of segments reconstructed along the U axis from $N$ images is given by :

$$P \ = \ N \ + \ K \ - 1 \tag{5.26}$$

Therefore, this reconstruction technique provides a way to estimate $P$ segments of the vertical aperture from $N$ measurements, the number of segments increasing with the distance to the sensor. However, it is important to note that the size of the segment is constant across ranges and equal to the sampling period along the vertical axis $\Delta_z$. As a consequence this method provides range-independent along-track resolution from any vertical aperture $A_v$.

## 5.5 3D reconstruction as a constrained optimization

As described in equation 5.23, the 3D reconstruction of the surfaces is equivalent to estimating $b$ given $I$ and $A$. The reconstruction of surfaces from SONAR measurements is a blind deconvolution problem (see section 5.2) since the coefficients $\tilde{\rho}^i_{jr\phi}$ depend on the shape of the object and therefore remain unknown. Therefore and in order to solve this linear system, our implementation includes additional assumptions and constraints.

In this section, we present the assumptions enabling the resolution of the sparse linear system formulation. We then discuss the choice of the regularization strategy and formulate the 3D reconstruction problem as an constrained MAP optimization. We then present two optional steps respectively adapted to operating reconstruction from noisy data and addressing deconvolution ringings.

### 5.5.1 Approximations

The elements in the matrix contain the coefficients $\tilde{\rho}^i_{jr\phi}$ which are dependent on the reflectivity distribution function of the object. Although, under the Lambertian assumptions, these terms would reduce to cosines, their calculation requires knowledge of surface normals which are *a-priori* unknown. We observed that ignoring them (setting them to unity) yields results that are comparable to more complex treatment. Therefore and similarly to the assumptions made in the space carving method (see section 4.2), we choose to assume the imaging model as view point independent and replace the $A$ matrix by an approximate model of the imaging process $A^\star$ such that :

$$A^\star_{ij} \;=\; \tilde{\beta}^i_j \tag{5.27}$$

### 5.5.2 Regularization

While regularization enables to constrain the solution of the problem, it needs to be adapted to the expected characteristics of the solution. Since the surface function to be reconstructed can typically be represented as a rectangular function, sparsity is desired to enforce a sparse solution and reduce typical deconvolution ringing and blurring artefacts (Liu and Jia [2008], Shan et al. [2008]).

Although $L_1$ regularization is commonly employed in optimization as a way to impose sparsity, recent study (Slawski et al. [2013]) suggests that applying a non-negativity constraint is sufficient to impose sparsity when the model matrix $A$ has certain properties which appear to be satisfied in the case of a typical deconvolution problem. In order to evaluate the interest of imposing sparsity we therefore choose to add a $L_1$ regularization constraint as an option.

### 5.5.3 Positivity constraint

In addition to regularization constraints, additional constraints can be expressed on the domain of the solution. In particular, a positivity constraint is often assumed in imaging problems. Following the convolution model presented in equation 5.2, the observed function is positive. We therefore formulate the 3D reconstruction problem as

an optimization problem with a positivity constraint:

$$\hat{b} \;=\; \underset{b \geq 0}{\arg\min} \;\; \|A^{\star}\, b - I\|_2 \;+\; \lambda \, \|b\|_1 \tag{5.28}$$

This $L_2$ optimization with positivity constraint is commonly referred to as a Non-Negative Least Square (NNLS) optimization problem.

### 5.5.4 Denoising and de-ringing

#### Z-test denoising of the measurements

In order to reconstruct surfaces from noisy data, we apply the same denoising technique as described in section 4.4.7, based on a Z-test and estimates of the first two moments $m(r, \phi)$ and $\sigma(r, \phi)$ obtained in open-water. The Z-test is then applied on the measurement vector $I$ to discard (set to zero) values lying in the noise distribution.

#### De-ringing of the solution

Since we solve for the discretized version $b$ of a surface indicator function $S$ across a slice in range (and bearing), the function we aim to reconstruct is typically composed of impulses where there is a surface. Consequently, the reconstruction of its discrete version via the above least squares formulation leads to severe ringing artifacts that cause 'halo' structures. To identify and eliminate these artefacts, we test segments in space where $b$ is non-zero to ensure that they were not observed to be 'free space' in any of the input images. Similarly to the space carving method presented in 4, if they were observed to contain no reflectivity in one unoccluded view, we set the corresponding element of $b$ to zero.

### 5.5.5 Implementation

The input to our algorithm is $N$ 2D SONAR images acquired using the same beam direction (bearing), the same $x$ and $y$ coordinates but from depths $z_i$, $i = 1, 2, ..., N$. We implemented the deconvolution reconstruction approach in C++ with a similar interface to our space carving implementation, allowing easy comparison between the two methods. Importantly, no 3D structure is needed for storage since each slice at $(r, \phi)$ is reconstructed independently from each other. The reconstruction is obtained using all the measurements, prohibiting an iterative reconstruction of the environment

by successive updates on the scene. The NNLS optimization is performed using an open-source solver Lawson and Hanson [1974]. In order to add $L_1$ regularization, a MATLAB implementation (Kwangmoo Koh and Boyd [2008]) of the truncated Newton interior-point method (Kim et al. [2007]) is used.

## 5.6 Results

### 5.6.1 Simulation

In this section, we present experimental results obtained with noise-free simulated data using the deconvolution methods without regularization with a vertical sampling period of 1cm. The same four reference models as the ones presented in section 4.5.1 have been inspected with two sensors of respective vertical apertures 1° and 10°.

Figure 5.5 presents reconstruction results obtained with both sensors on noise-free data. Conversely, pencil-beam reconstructions exhibit finer reconstructions on small parts of the structures such as the wheel valve in figure 5.5-e and the side tubes on figure 5.5-k.

In order to visually compare the deconvolution reconstruction to the space carving technique, we present in figure 5.6 the point clouds obtained with both methods and both sensors.

#### 5.6.1.1 Quantitative results of noise-free reconstructions

As performed with the space carving method, we present here a quantitative analysis with various vertical sampling period. In order to evaluate the performance of the deconvolution method and compare it to the space carving result, we based our analysis on the same four metrics presented in section 4.5.2.1.

Similarly to the carving reconstruction results, the median error of the deconvolution reconstruction (see figure 5.7) increases with larger sampling periods when using a wide-beam SONAR. On the contrary, when using a pencil-beam sensor, the median reconstruction error does not appear to depend on the sampling period with constant error values under 1cm.

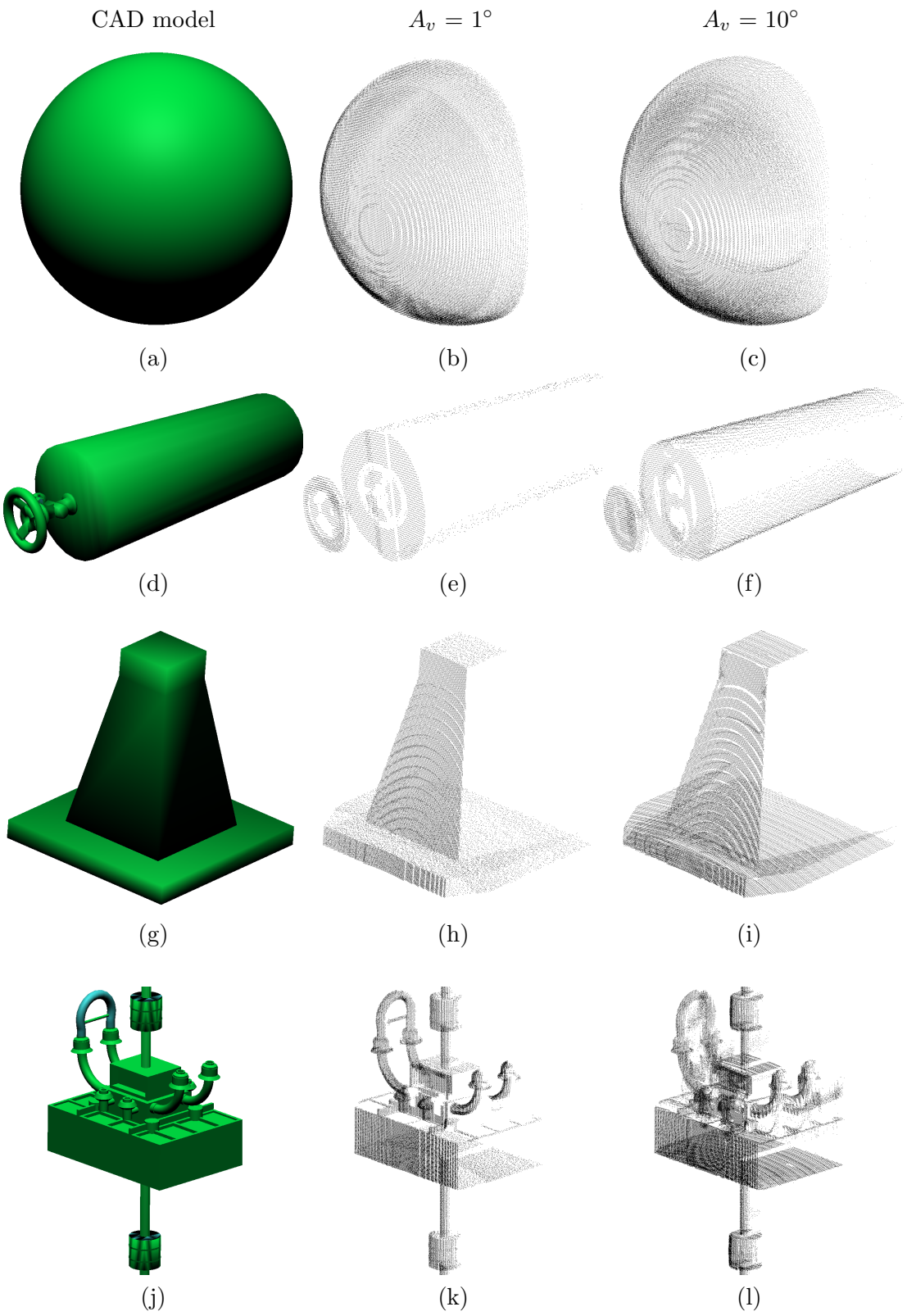



FIGURE 5.5: 3D deconvolution reconstruction from noise-free simulated data obtained after inspection of the four reference models. CAD models (a,d,g,j), pencil-beam SONAR reconstructions (b,e,h,k) and wide-beam sensor reconstructions (c,f,i,l).

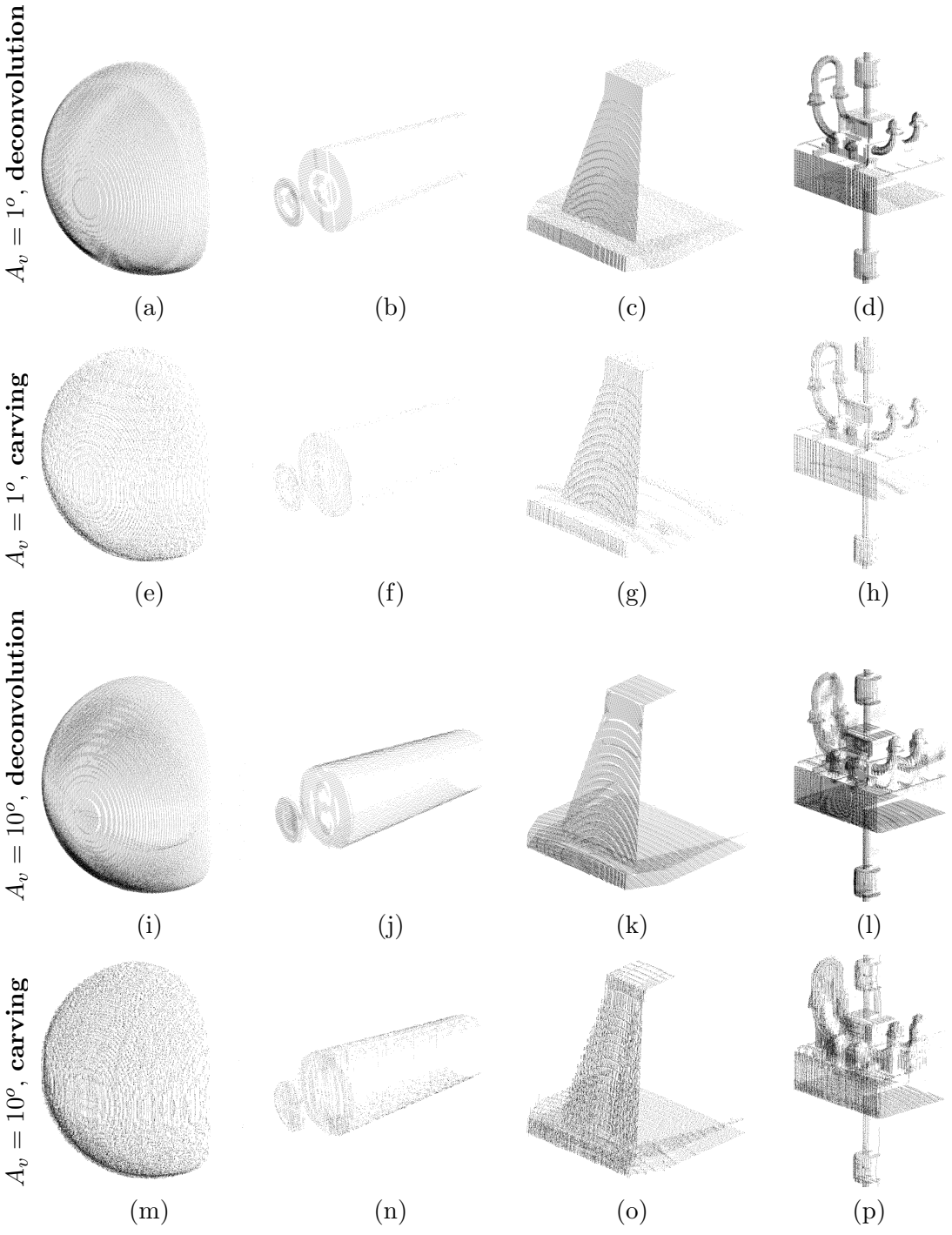



FIGURE 5.6: 3D reconstruction results from simulated SONAR images using our two algorithms with two sensors of different vertical aperture: Pencil-beam reconstructions (a-h) and wide-beam reconstructions (i-p). Our algorithms reconstructed point clouds using the deconvolution method (a-d and i-l) and the space carving technique (e-h and m-p).

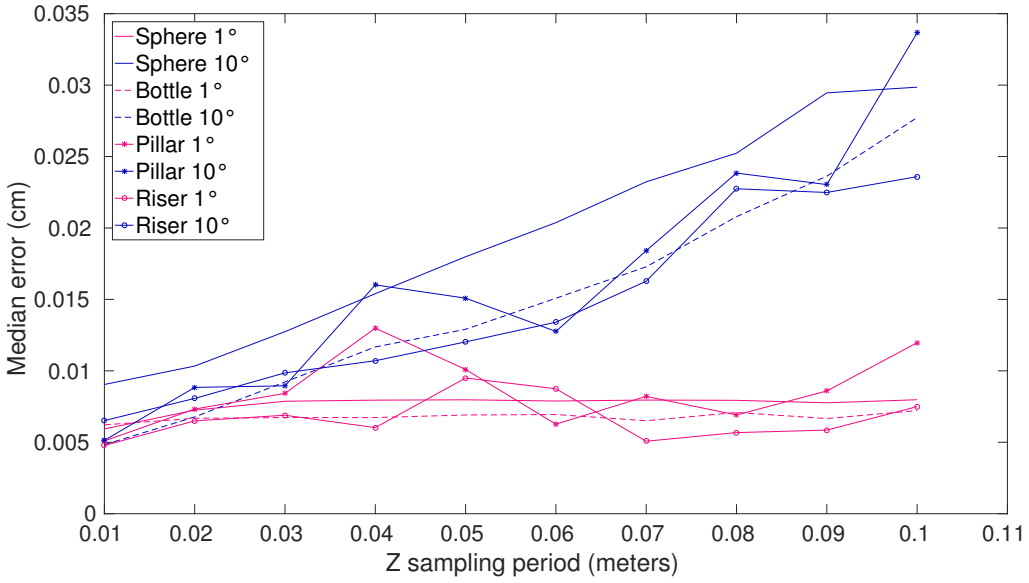



FIGURE 5.7: Median error of the deconvolution reconstructions versus vertical sampling period on all models with two different sensors: pencil-beam (magenta) and wide-beam (blue). Similar reconstruction errors are obtained when considering small sampling periods but the error when using a pencil-beam sensor does not appear to depend on the sampling period.

As visible in figure 5.8, the surface coverage decreases monotonically with larger sampling periods. Similarly to the carving results and in comparison to when using a pencil-beam SONAR, surface coverage values are roughly twice as high when using a wide-beam sensor except on the sphere structure. As can be observed in figure 5.9, when combining the two previous metrics in a single median error to coverage ratio, very similar results are obtained with the two sensors at each vertical sampling period. The increase in coverage obtained when using a wide-beam sensor is balanced by a decrease in accuracy in comparison to pencil-beam SONARs. As illustrated in figure (5.10), pencil-beam reconstructions are outlier-free whereas reconstructions from wide-beam sensors generate a small proportion of outliers with increasing presence at larger sampling periods.

We present in figure 5.11 a comparative analysis between the deconvolution and the space carving reconstruction technique with both a pencil-beam and a wide-beam sensor at multiple sampling periods. In order to compare the two algorithms, the four reference metrics were averaged over the multiple structures used in simulation. Figure 5.11-a shows similar reconstruction accuracy using both methods while 5.11-b exhibits the increase in surface coverage when using the deconvolution method over the space carving method. As a consequence, the median error to coverage metric shows lower values when

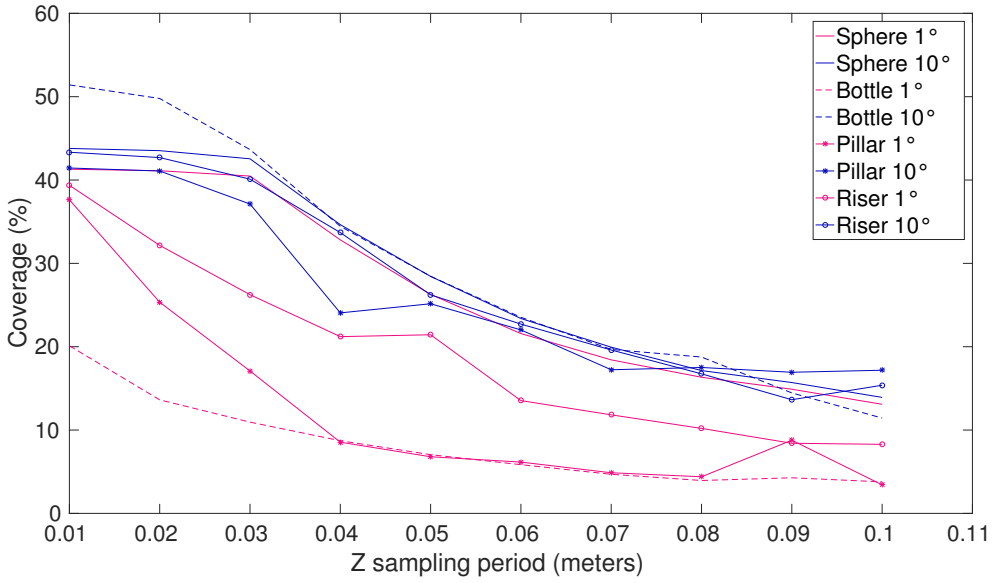



FIGURE 5.8: Surface coverage of the deconvolution reconstructions versus vertical sampling period. The surface coverage yielded when using a pencil-beam sensor is twice as low as when using a wide-beam sensor except on the sphere object where a similar coverages are observed.

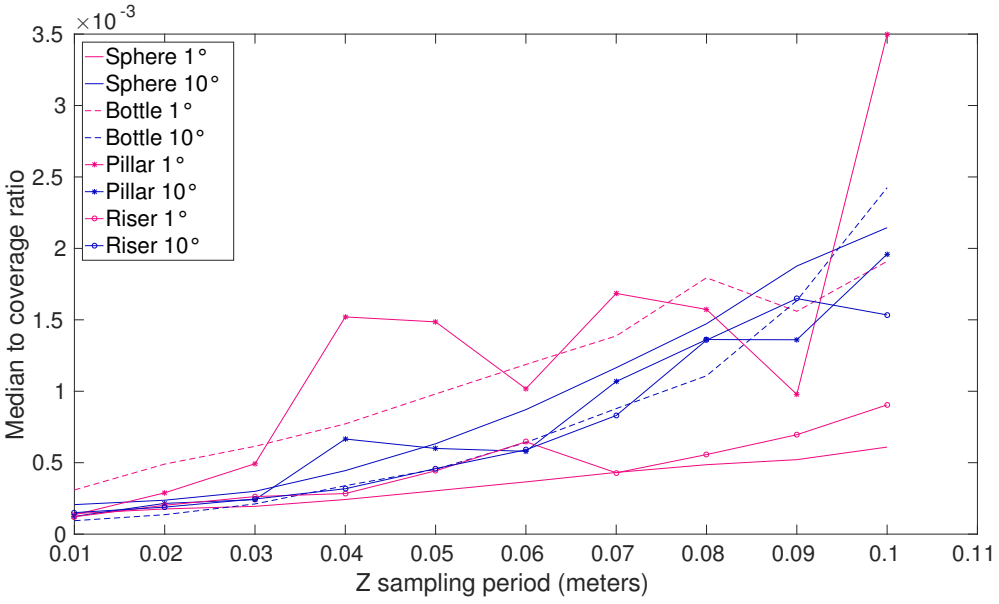



FIGURE 5.9: Median error to coverage ratio of the deconvolution reconstructions versus vertical sampling period. The reconstructions obtained with the two sensors exhibit similar median to coverage values.

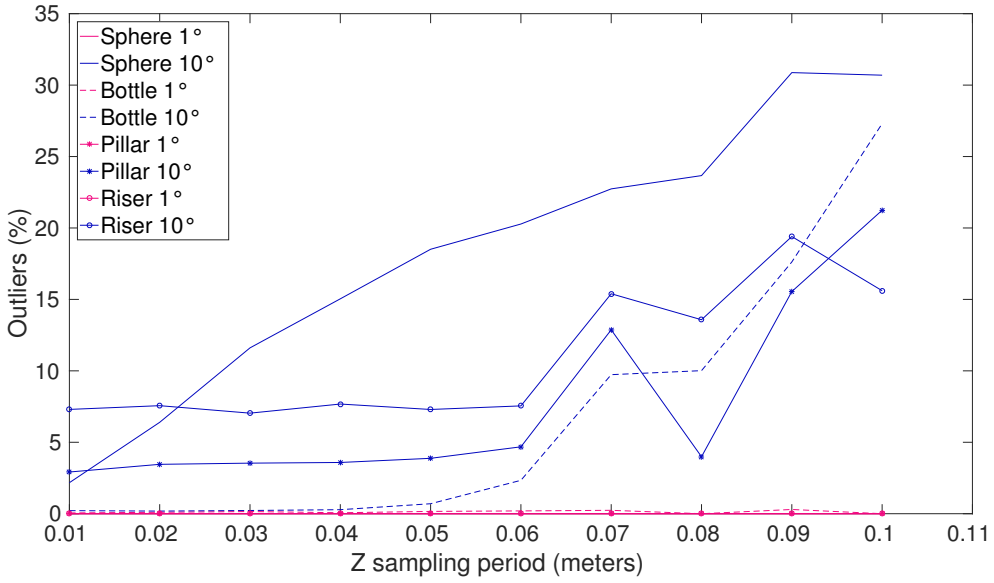



FIGURE 5.10: Presence of outliers in the deconvolution reconstructions versus vertical sampling period. While no outliers appear in pencil-beam reconstructions, a small proportion of outliers is observed in wide-beam reconstructions, particularly when considering large vertical sampling periods.

using the deconvolution method. Figure 5.11-d shows similar outliers rates for both reconstruction methods.

**Importance of satisfying the rectilinear approximation**

We present here quantitative results obtained when considering larger vertical apertures than imposed by the rectilinear approximation. Figure 5.12 shows median errors and coverage values for sensors of respective vertical apertures 1°, 10°, 20°and 30°. Figure 5.13 represents the error to coverage values obtained with the four sensors. As illustrated in figure 5.14, when the conditions of observation do not enable the rectilinear approximation to remain valid, significant artefacts can be observed on vertical surfaces.

#### 5.6.1.2 Analysis

Qualitatively, the deconvolution method exhibit a good level of geometrical details with a higher coverage of horizontal surfaces when using the wide-beam sensor (figure 5.5-f) in comparison to pencil-beam reconstructions (figure 5.5-e). The visual comparison of these results showed the impossibility of observing horizontal surfaces with a pencil-beam SONAR (e) while the equivalent wide-beam reconstruction (f) exhibits increased

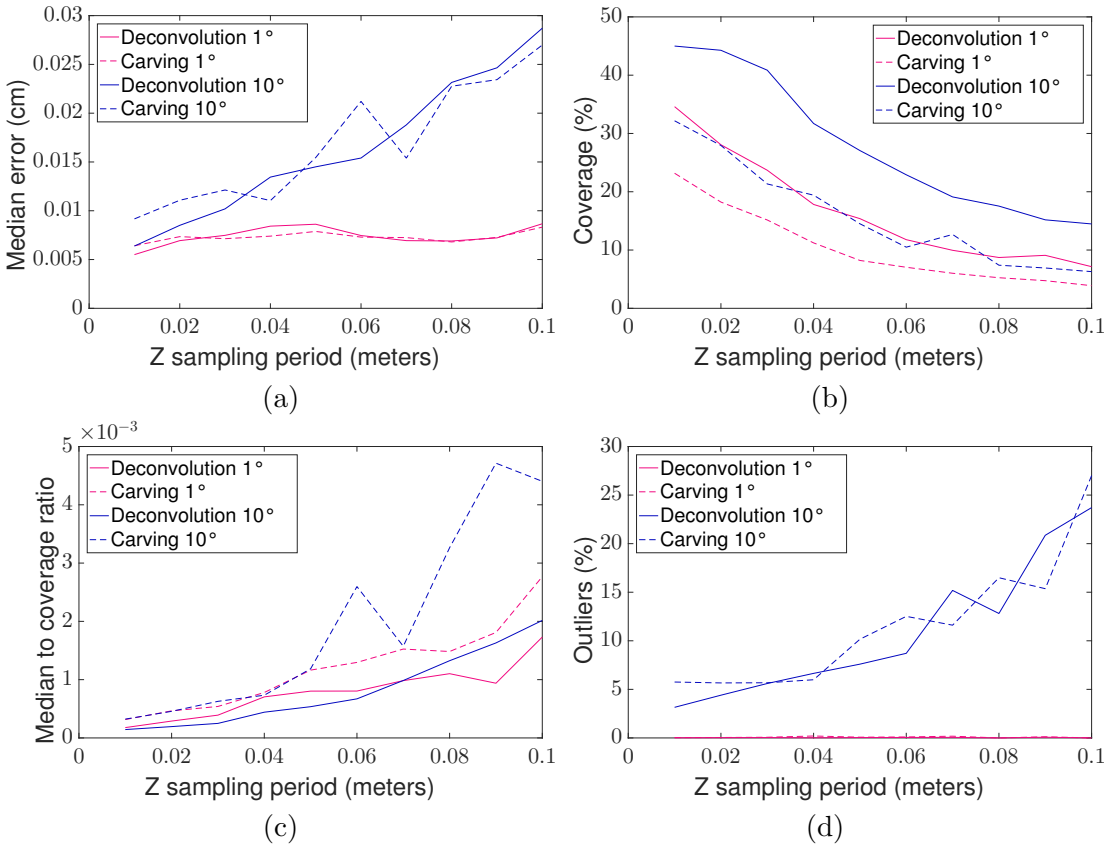



FIGURE 5.11: Averaged quantitative results of deconvolution vs space carving simulated reconstructions. The four reference metrics averaged over all structures are presented for both algorithms (different pattern) and both sensors (different colours). While similar median errors are obtained with both reconstruction techniques (a), the deconvolution offers significantly higher surface coverage (b) with both the pencil-beam and the wide-beam sensor. As a result, the combined median error to coverage metric (c) appears consistently lower using the deconvolution method. Outlier rates appear similar with both algorithms (d).

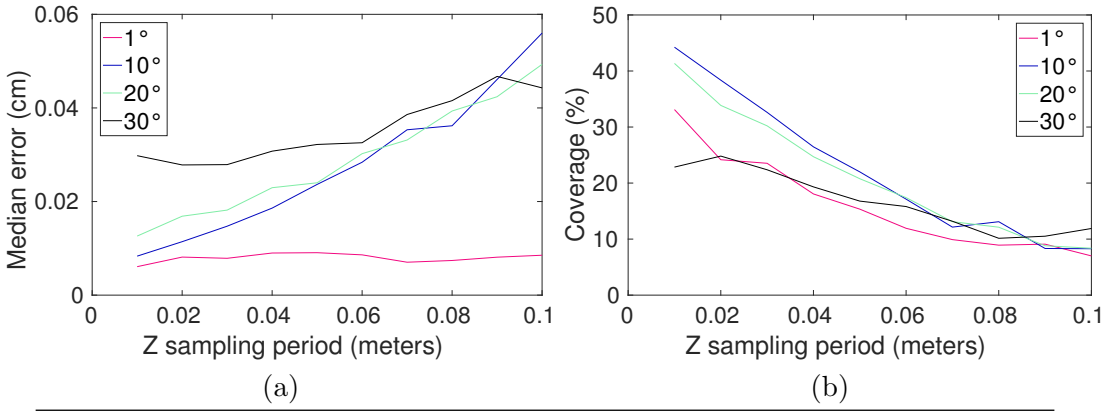



FIGURE 5.12: (a) Median error versus vertical sampling period averaged on all models for different sensor apertures. The reconstruction error when using a 30° sensor appears consistently higher than when considering smaller apertures. (b) Coverage versus vertical sampling period averaged on all models for different sensor apertures. The coverage value when using a 30° sensor appears just as low as when using a 1° sensor.

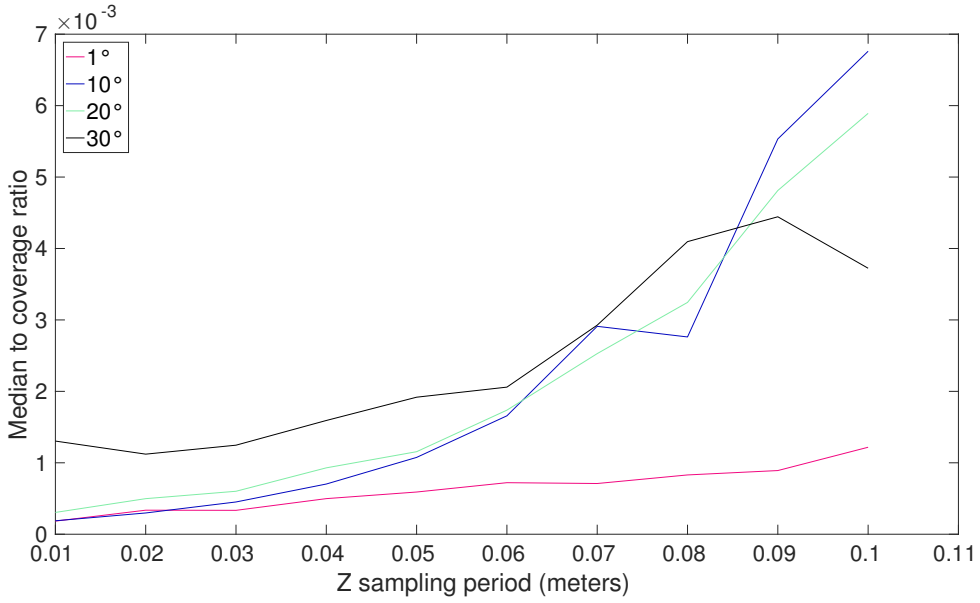



FIGURE 5.13: Median error to coverage ratio averaged on all CAD models versus vertical sampling period for different sensor vertical apertures. Applying the deconvolution method to a 30°vertical aperture sensor provides significantly higher error rates in comparison to smaller apertures, in particular at low vertical sampling period (up to 5cm).

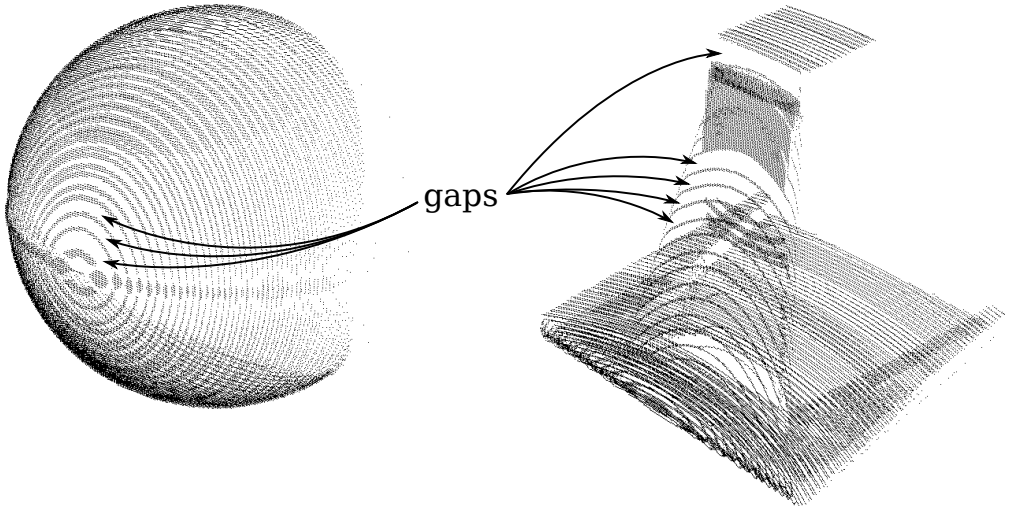



FIGURE 5.14: Illustration of a degraded deconvolution reconstruction due to the rectilinear approximation: when using a 30°vertical aperture sensor and a 0.01cm sampling resolution at more than 2 metres range, the rectilinear approximation does not remain valid and generates significant gaps on vertical (rectilinear) surfaces.

coverage but lower geometrical details as visible on the wheel valve. Similarly when comparing the two reconstructions methods in figure 5.6, the point clouds obtained with the deconvolution method exhibit better regularity (see sphere figures (a,i) vs figures (e,m)) resulting in an overall denser appearance. Horizontal surfaces appear better covered as visible on the pillar figures (c,k) compared to figures (g,o). The edges of the structures also appear sharper as can be observed in the riser structure (figures (d,l) vs (h,p)).

The quantitative results showed the consistent higher surface coverage obtained when using the deconvolution method. Similar median errors were obtained with the space carving method resulting in overall better 3D reconstructions when using the deconvolution technique. Regarding the difference in performance when using different sensors, figure 5.11-a and figure 5.11-b confirmed the results previously obtained with the separated analysis of the two algorithms exhibiting the improvement in surface coverage when using a wide-beam sensor at the cost of a decrease in reconstruction accuracy. Figure 5.11-c offered an interesting perspective on the choice of the sensor for 3D reconstruction by exhibiting better overall reconstruction metric when using a wide-beam sensor and the deconvolution reconstruction technique for sampling rates of up to 7cm. Not only this counter-intuitive result showed the good performance of the deconvolution algorithm by solving for vertical uncertainty but it also justifies the use of wide-aperture sensors for 3D reconstruction by levering their larger footprint at the cost of small decrease in accuracy with a small proportion of outliers (figure 5.11-d). Finally and as could be expected from the rectilinear assumption, figure 5.13 showed that the deconvolution method does not perform well when considering larger vertical apertures ($\geq 22°$ in the case of our test scenario). The employment of this method should therefore be conditioned by the satisfaction of the rectilinear assumption which in turn will impose a maximum vertical aperture or conversely, shorter measurement range or lower range resolution.

### 5.6.2 Tank experiment

Since the experiment made in Heriot Watt University water tank (see section 4.6.1 for description) was obtained using a rectilinear motion of the sensor with uniform sampling, the deconvolution reconstruction techniques is applicable directly to the dataset with both the BlueView MB2250 and the Aris Explorer 3000.

We present the results of 3D reconstruction using the deconvolution method on the tank dataset in figure 5.15. Figures 5.15-a,c,e,g present the reconstructions obtained from the pencil-beam SONAR (BlueView MB2250) whereas figures 5.15-b,d,f,h show the equivalent reconstructions obtained with a wide-beam SONAR (Aris Explorer 3000). In order to visualize the difference between the deconvolution results and the space carving results, figure 5.16 compares the tank reconstruction results with both sensors and two structures: the sphere and the Hyball ROV.

#### 5.6.2.1 Analysis

The accuracy of the 3D reconstruction using the deconvolution method is similar or better than the space carving. In particular, the reconstruction of the sphere obtained with the Aris Explorer (figure 5.15-b) exhibits greater detail than its equivalent with the BlueView MB2250 with a very clear rope laying on top and side of the sphere. The pencil-beam reconstruction of the sphere exhibits good shape coherency but suffers from holes in the 3D point cloud which would result in a lower surface coverage. A similar observation can be made on the Hyball ROV example (figure 5.15-g-h), where the wide-beam reconstruction appear more complete than their pencil-beam equivalent. While the cylinder structure does not offer many features, the rope can again be noticed in the wide-beam reconstruction which is not the case in the pencil-beam reconstruction. Although the wide-beam deconvolution reconstructions exhibit a good level of details, a greater amount of outliers can be observed, cluttering the representation (see figure 5.15-f).

The differences between the two reconstruction methods can be visualised in figure 5.16: when using the pencil-beam, the deconvolution reconstructions appear fuller and the wide-beam deconvolution reconstructions exhibit more detail with sharper edges than their space carving equivalent. The sphere and rope reconstruction (figure 5.16-e) is thus sharper than its carving reconstruction equivalent (figure 5.16-g). In accordance to the simulation result, the difference between the two reconstruction techniques is more noticeable when using a wide-beam SONAR. Although the deconvolution reconstruction generate a small proportion of outliers, these latter appear sparsely as opposed to the space carving reconstruction results where outliers tend to be grouped in blocks.

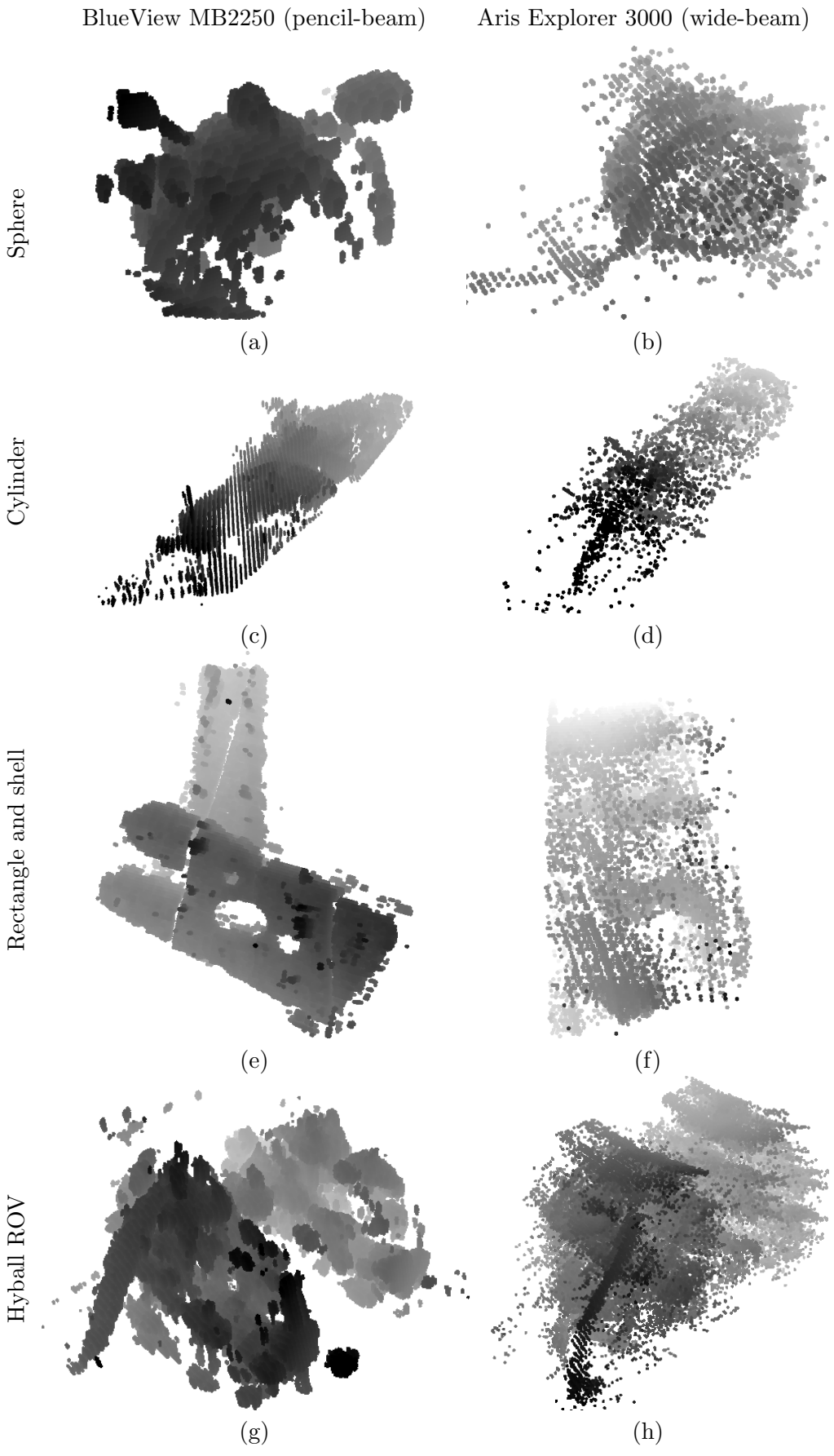



FIGURE 5.15: 3D deconvolution reconstruction from tank data with several reference objects (rows) and two sensors of different vertical aperture (columns).

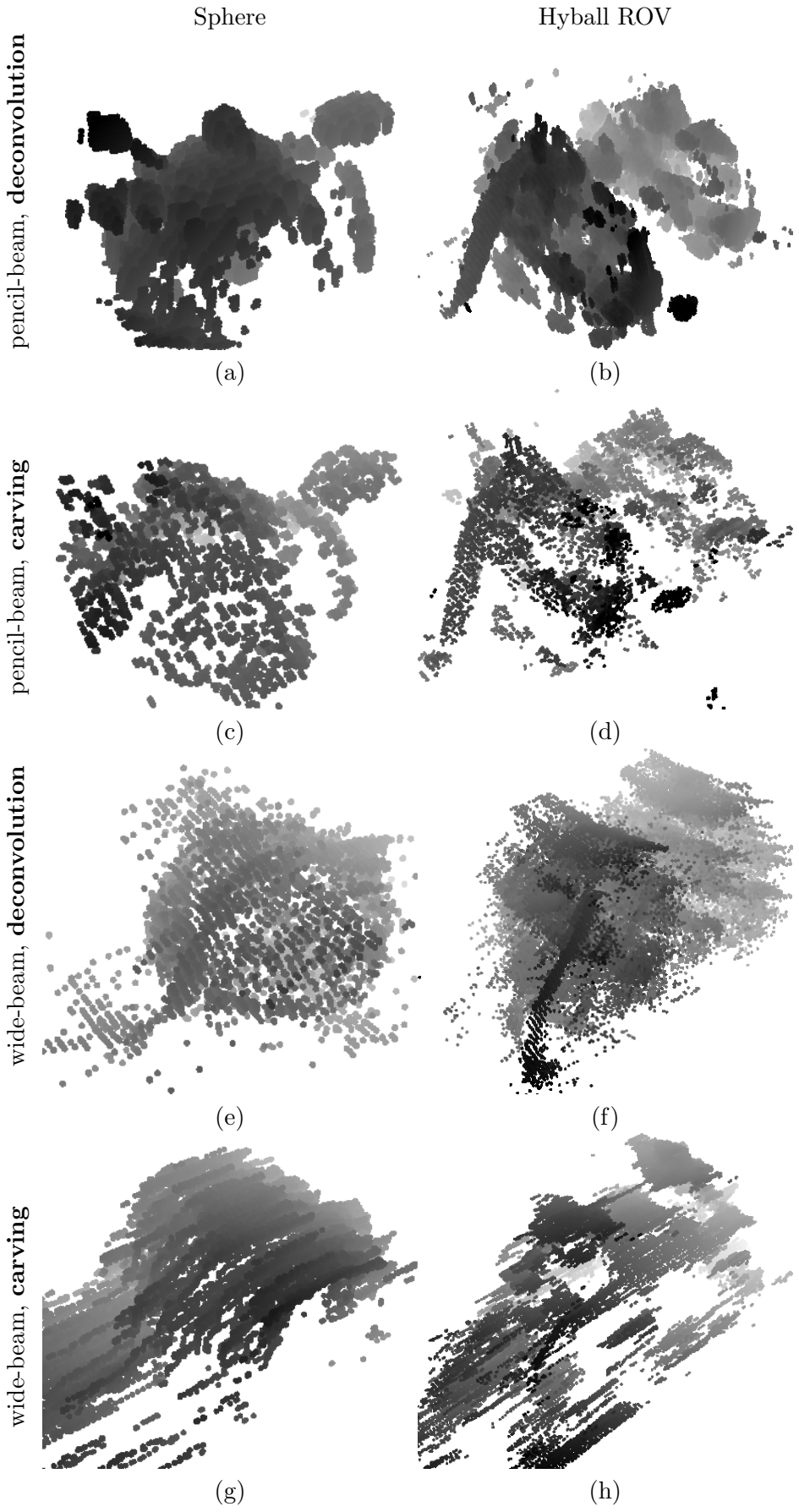



FIGURE 5.16: Comparison of 3D reconstruction from tank data with deconvolution and space carving methods. Two objects are being compared: the sphere (first column) and the Hyball ROV (second column) with both pencil-beam (a-d) and wide-beam sensors (e-h).

### 5.6.3 Field data

We present here reconstruction results obtained with field data gathered by the AIV prototype (see section 4.7.1). The vehicle performed inspections using two different acoustic sensors, a downward-looking pencil-beam BlueView MB2250 (1° vertical aperture) and a forward-looking wide-beam BlueView P900-130 (20° vertical aperture). While the wide-beam sensor featured a large enough footprint for the deconvolution method to apply (significant overlap with $K \geq 2$), pencil-beam images were generally acquired at low sampling rate (4cm) leading to a small overlap between each consecutive samples.

#### 5.6.3.1 Lake trials

The first field dataset was acquired during AIV lake trials (see description in section 4.7.2).

**Pencil-beam reconstruction**

We present in figure 5.17 an example of deconvolution reconstruction using the pencil-beam sensor with a small overlap ($2 \leq K \leq 3$) obtained while performing a top inspection of the box structure. Due to the small overlap between consecutive samples


(a) CAD model (b) Deconvolution (c) Carving

FIGURE 5.17: Comparison of 3D deconvolution reconstruction (b) of the box structure versus its equivalent carving reconstruction (c) using pencil-beam data (BlueView MB2250). Both points clouds are represented as depth map. a) Reference CAD model of the inspected structure.

(small aperture and low sampling rate), the two reconstructions are very similar. Unlike the space carving where points are located following the imaging model and the position of the sensor at the time of the sampling, the deconvolution method assume a rectilinear and uniform motion leading to very clear scanning lines in the reconstructed point cloud.

**Wide-beam reconstruction**

In addition to pencil-beam reconstructions, we applied the deconvolution reconstruction technique to the field dataset gathered by the AIV project using the wide-beam BlueView P900-130. Although not strictly rectilinear (± 2cm) and not strictly uniform (± 2cm), the meter long vertical movement of the vehicle while inspecting the box structure was approximated by a vertical and uniform motion in order to satisfy the deconvolution model.

Figure 5.18 shows the deconvolution reconstruction of the box structure (5.18-b) in comparison to its CAD model (5.18-a) and the space carving reconstruction (5.18-c). In comparison to the space carving reconstruction, the deconvolution reconstruction


(a) CAD model (b) Deconvolution (c) Carving

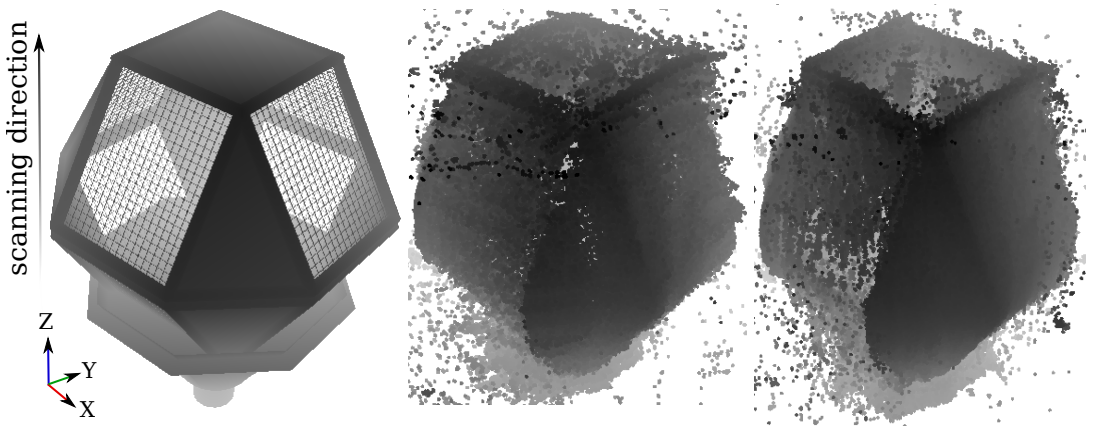
FIGURE 5.18: Comparison of 3D deconvolution reconstruction (b) of the box structure versus its equivalent carving reconstruction (c) using wide-beam data (BlueView P900-130). Both points clouds are represented as depth map. a) Reference CAD model of the inspected structure.

appears to be more regular and features both the outer box and the inner box. The carving reconstruction does not exhibit the full inner box and appears more cluttered. As observed previously in the tank experimental results, the outliers structure differs between the two reconstruction techniques: although very visible, the outliers present in the deconvolution reconstruction appears to be more spread than the outliers in the carving reconstruction as clustered in groups of vertical orientation.

#### 5.6.3.2 Offshore trials

The 360° offshore field inspection data around the polygonal structure described in section 4.7.3 was reconstructed to evaluate the deconvolution algorithm performance on a large structure (8m high) observed at low sampling rate. As visible in figure 5.19-b,

the deconvolution reconstruction is dense and appears coherent with the CAD model (5.19-a). In comparison, the space carving reconstruction (c) features a little less outliers




(a) CAD model (b) Deconvolution (c) Carving

FIGURE 5.19: Comparison of 3D deconvolution reconstruction (b) of the polygonal structure versus its equivalent carving reconstruction (c). a) Reference CAD model of the inspected structure.

but a few points are missing on one face of the structure while this is not the case in the deconvolution point cloud. Similarly to the observations made with the tank and lake experiments, the outliers spatial distributions differ in the two point clouds. A larger number of outliers is observed at the extremities of the point cloud (bottom and top) in the deconvolution reconstruction while the outliers in the space carving point cloud appear more uniformly spread along the Z-direction.

### 5.6.4 Influence of regularization

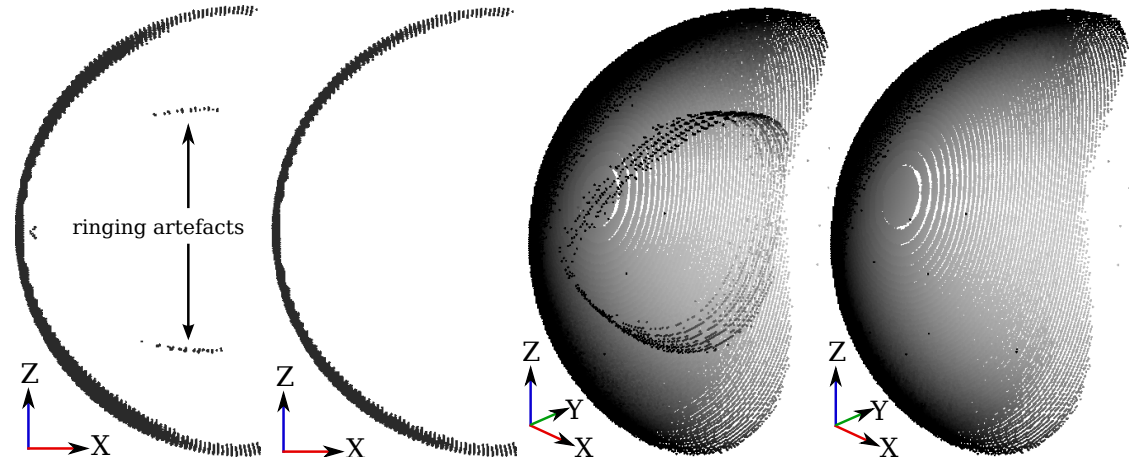
In this section, we study the influence of added regularization on 3D reconstruction results using the deconvolution technique. Following the description of results on both simulated and real data, we highlight the practical limitations of regularization.

#### 5.6.4.1 Regularized reconstruction

The choice of the regularization parameter $\lambda$ balances the importance given to the data fidelity (relatively to the imaging model) versus the regularity of the solution. We present here results obtained with various values of $\lambda$, expressed as a fraction of $\lambda_{max}$ where $\lambda_{max}$ is the maximum value of lambda that make the solution non-zeros (defined in Kwangmoo Koh and Boyd [2008]).

### Simulated data

We present here a regularized reconstruction of the sphere structure and focus in particular on the inside of the object. As illustrated in figure 5.20, adding regularization to the optimization process enables to impose sparsity. When imposing sparsity, not only the




(a) No regularization (b) $\lambda = 0.5\ \lambda_{max}$ (c) No regularization (d) $\lambda = 0.5\ \lambda_{max}$

FIGURE 5.20: Illustration of the influence of the $L_1$ regularization on noise-free simulated data. Figures (a,b) show a vertical slice ($x = 0$ plane) of the reconstructed sphere while figures (c,d) show the inside of the fully reconstructed sphere (back view). In comparison to unregularized reconstructions (a,c), the surface of the regularized reconstructions (b,d) appears to be thinner. In addition to a thinner surface, the regularization removes the deconvolution ringing artefacts observed inside the spheres in figures (a) and (c).

regularization provides a thinner object surface (as visible in figures 5.20-a,b with the vertical slices of the object) but it also enables to suppress the deconvolution ringing artefacts (see depth maps of the inside of the sphere in figures 5.20-c,d) as frequently observed in the presence of sharp edges (pulse along the Z-direction). As detailed in table 5.1, the regularization improves the reconstruction error by 30% without changing the coverage. As could be expected from the visual analysis, the regularization also reduces

| | **No regularization** | **$L_1$ regularization** |
|---|---|---|
| **Median error (cm)** | 0.9 | 0.65 |
| **Coverage (%)** | 43.79 | 43.77 |
| **Outliers (%)** | 2.164 | 0.019 |

TABLE 5.1: Influence of $L_1$ regularization ($\lambda = 0.5\ \lambda_{max}$) on the reconstruction of a sphere.

the proportion of outliers by suppressing the ringing artefacts.

### Tank data

As illustrated in figure 5.21, we applied the regularized deconvolution algorithm to the

aluminium sphere dataset gathered in the water tank provides a reconstruction of the sphere that appears more dense (top surface of the object more densely represented) and the object size varies slightly (5%) along the scanning direction.

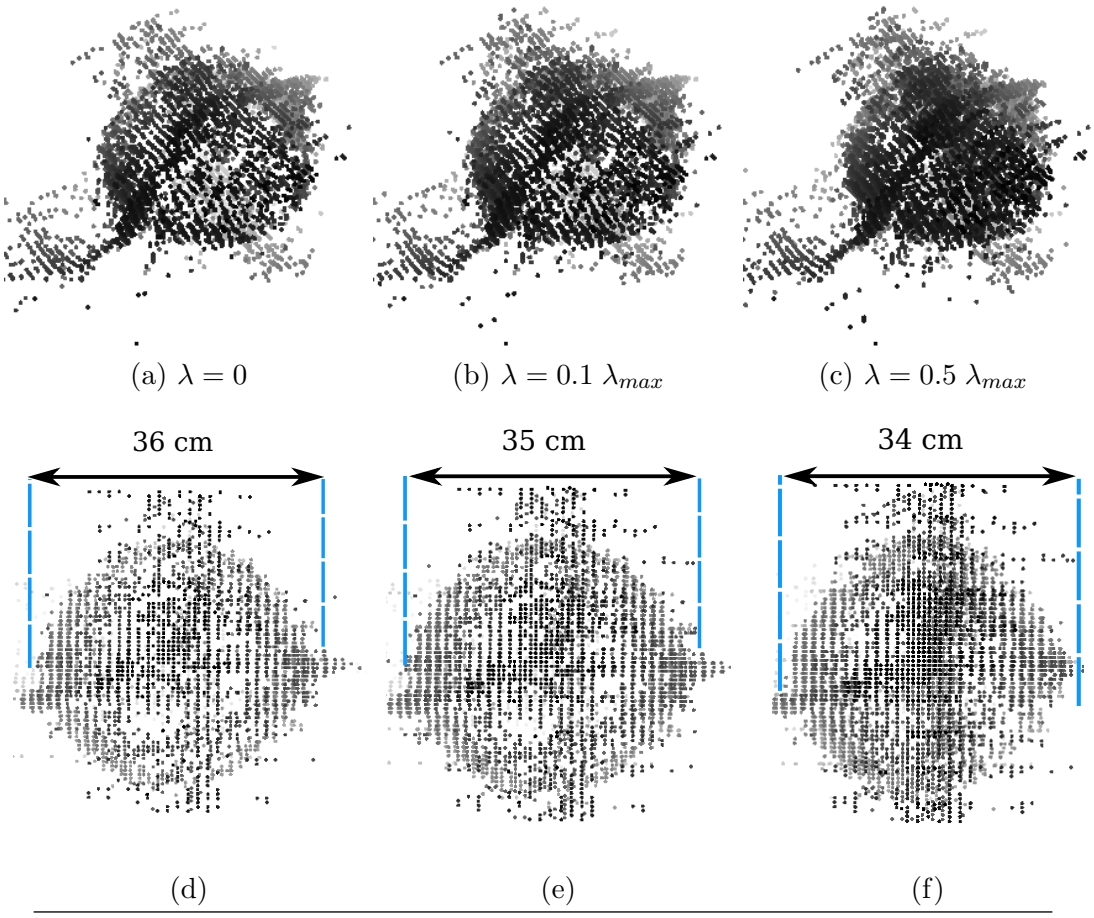


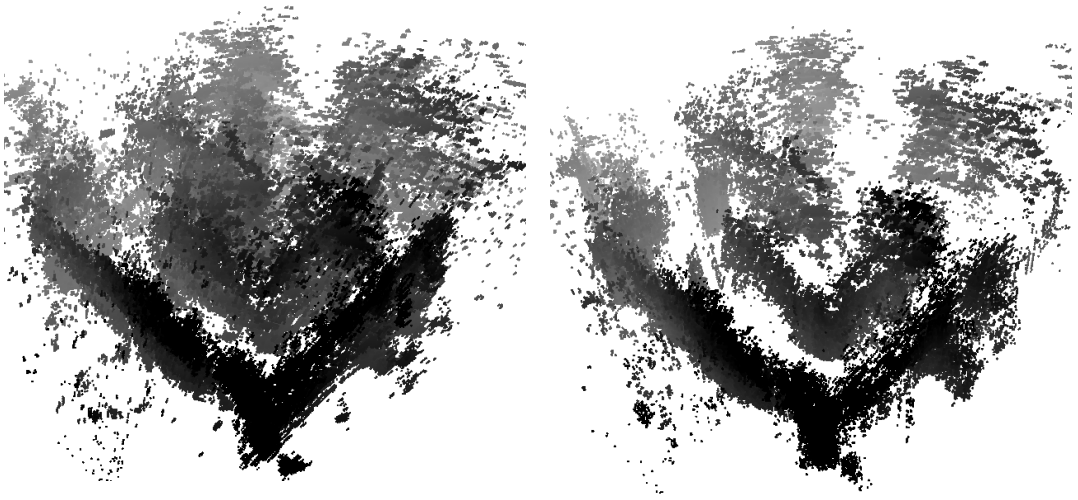

FIGURE 5.21: Illustration of the influence of the $L_1$ regularization term. In comparison to the regularization-free reconstruction (a), the mild regularization (b) appears more dense. Using a strong regularization term (c) enables an even more dense reconstruction making the object size smaller along the scanning direction. The regularization also has a clustering effect as observed in the top views of the reconstructions (d,e,f).

**Field data**

As illustrated in figure 5.22, applying $L_1$ regularization on the reconstruction of the wide-beam dataset of the box structure leads to a sparser 3D representation.

#### 5.6.4.2 Limitations

While $L_1$ regularization is commonly employed in optimization problems as a way to impose sparsity, recent study (Slawski et al. [2013]) suggests that applying a non-negativity constraint is sufficient to promote sparsity when the model matrix $A$ has



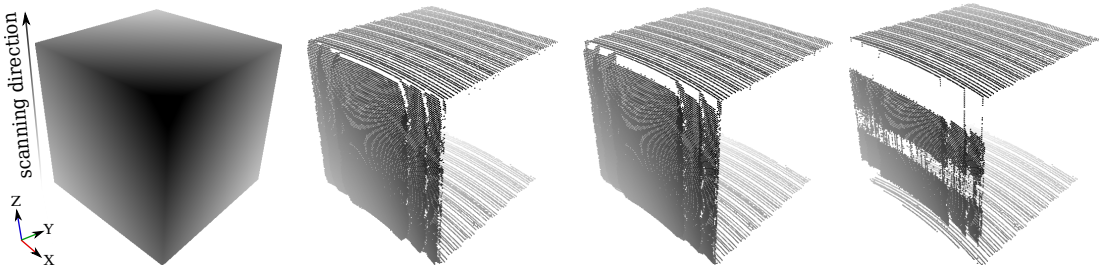
(a) No regularization (b) $\lambda = 0.1\ \lambda_{max}$

FIGURE 5.22: Illustration of the influence of the $L_1$ regularization term. In comparison to the unregularized reconstruction (a), the regularized reconstruction (b) appears more sparse.

certain properties which appear to be satisfied in the case of a typical deconvolution problem. Our experiments showed the interest of using $L_1$ regularization to impose further sparsity but also highlighted the limitations of regularization.

**Reconstruction of smooth surfaces along the sampling direction**

Although sparsity promotion is desired when reconstructing rectangular functions as occurring when observing a surface in a non-orthogonal way, surfaces that appear continuous along the sampling direction should be reconstructed as smooth sections rather than sparse sets of points. This situation is illustrated in figure 5.23 where gaps appear on the front surface of the reconstructed cube when using $L_1$ regularization due to the enforcement of sparsity on a solution that should appear as continuous. In this




(a) CAD model (b) No regularization (c) $\lambda = 0.1\ \lambda_{max}$ (d) $\lambda = 0.5\ \lambda_{max}$

FIGURE 5.23: Limitations of sparsity promotion with $L_1$ regularization. In comparison to the CAD model (a), the unregularized reconstruction (b) appears complete while the $L_1$ regularized deconvolution (c,d) exhibits gaps on the front face of the cube due to sparsity promotion.

situation, unregularized deconvolution (figure 5.23-b) provides an accurate reconstruction of the object. $L_1$ regularization is therefore not suited for reconstructing a planar section observed perpendicularly such as a wall or a flat seabed. As a consequence, $L_1$ regularization can only be applied in specific situations.

**Total variation regularization**

In order to obtain a regularization formulation that promotes both sharp edges (sparsity) and smoothness (continuity), the total variation regularization term (Rudin et al. [1992]) can be employed. By effectively imposing sparsity on the derivative of the function along the direction of deconvolution, the total variation approach typically reduces the number of transitions in the reconstructed signal, resulting in a formulation adapted to the description of quantized signals. In practice, our experiments on simulated data with a total variation regularization strategy highlighted the difficulty of the choice of the regularization parameter. In particular, the simulated sphere, featuring a wide range of surface normals provided range-dependent PSF (Point Spread Functions) for which no common regularization parameter was found.

**Regularization parameter choice**

In general, experiments made with TV and $L_1$ regularization showed that the choice of the regularization parameter $\lambda$ was not straight forward and appears to depend on the level of noise in the input data as well as the PSF. As a consequence, $\lambda$ is obtained empirically after several iterations with different values.

### 5.6.5 De-ringing and denoising

In addition to regularization, two optional processing steps can be added to improve the quality of the reconstruction. As described in section 5.5.4, a denoising step based on a previously established sensor-specific noise model is applied directly on the input SONAR images before the reconstruction process. Similarly a de-ringing step is applied after reconstruction to suppress ringing artefacts by comparing the reconstructed values to open water observations in the input SONAR images. In order to illustrate the importance of these two steps, figure 5.24 shows reconstruction results on the aluminium sphere dataset with and without deringing or denoising. As visible in figure 5.24-b, visible ringing artefacts remain when no deringing is performed. Figure 5.24-d shows the importance of

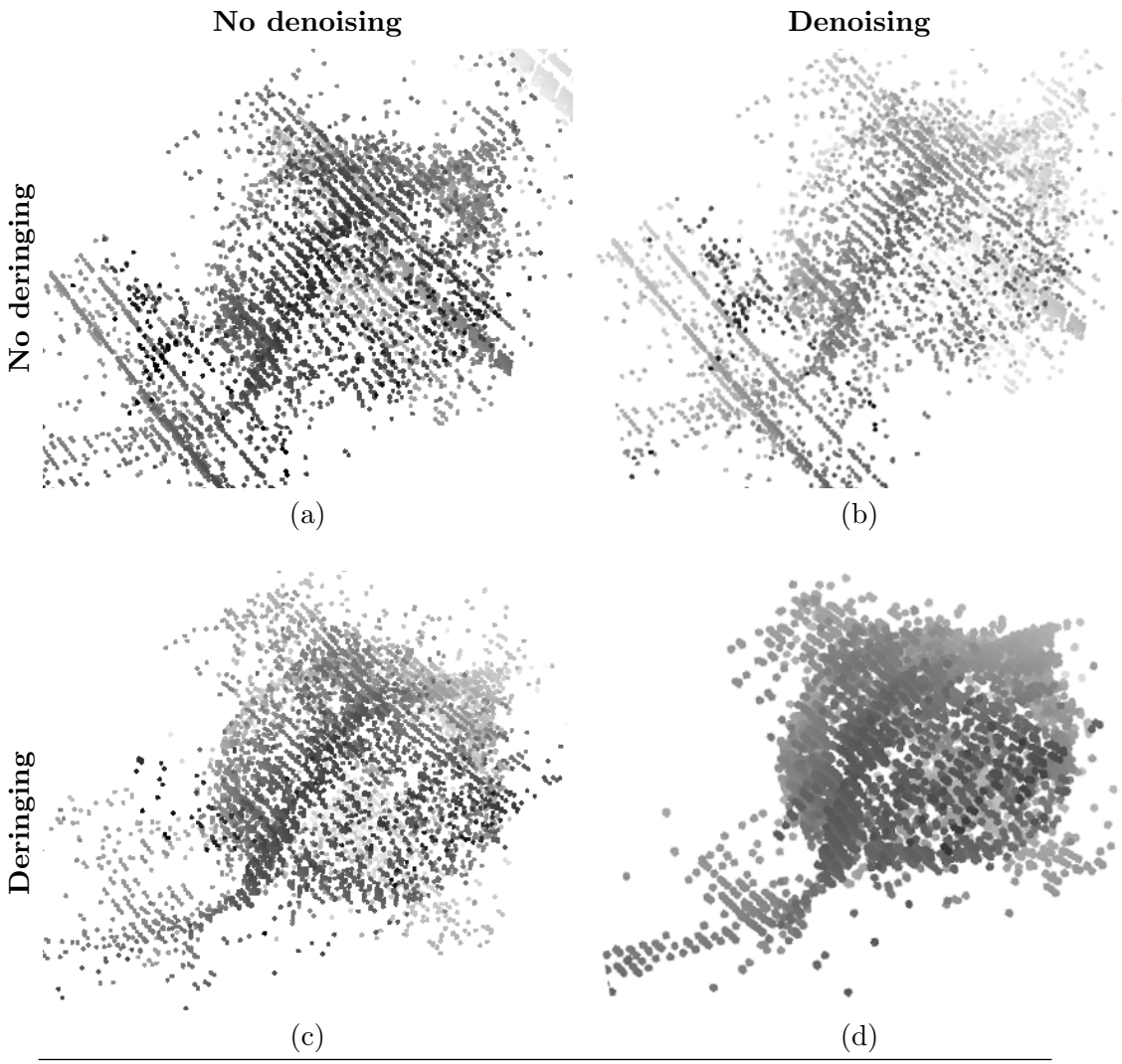
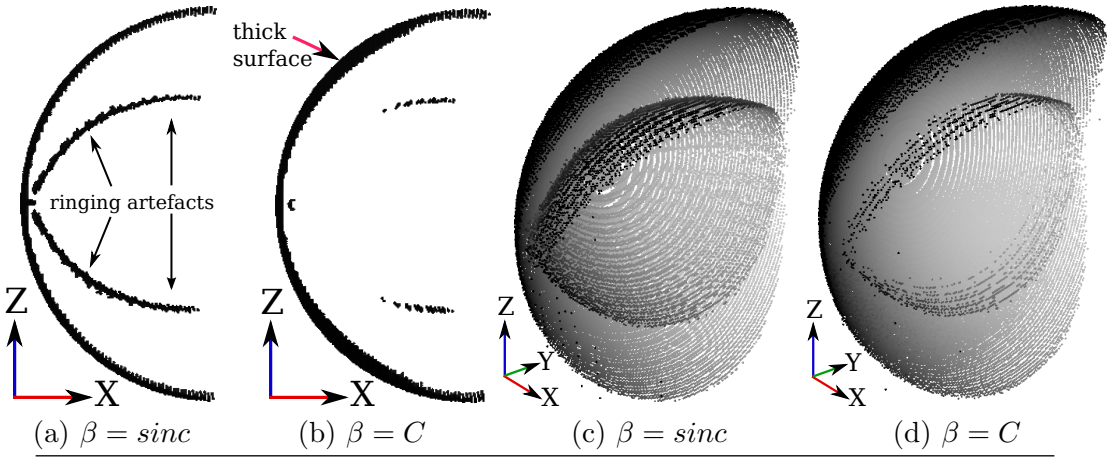



FIGURE 5.24: Illustration of the importance of de-ringing and denoising steps.

combined deringing and denoising in comparison to the original reconstruction (figure 5.24-a).

### 5.6.6 Beam pattern thresholding

The function $\beta$ ideally contains a step profile in the Fourier domain (low-pass filter). As a consequence, the typical beam patterns used for the sensors are Bessel functions. While these patterns are known a priori, for any given sensor, we observed that they do not play a major role in the approximation. We observed that providing no beam-pattern adjustment produced lower ringing artefacts than when quantized approximations of the beam pattern were used (see figure 5.25). As visible in this example, considering a flat beam pattern ($\beta = C$) reduces the amount of ringing artefacts inside the objects at the




(a) $\beta = sinc$ (b) $\beta = C$ (c) $\beta = sinc$ (d) $\beta = C$

FIGURE 5.25: Influence of the beam pattern profile. While the profile of the ideal beam pattern $\beta$ follows a Bessel function (sinc), experiments were made with a flat beam pattern assumption during the deconvolution (b,d) reducing ringing artefacts while increasing the width of the reconstructed surface. Deconvolution results obtained with a Bessel profile assumption (a,c) exhibit strong ringing artefacts inside the object.

cost of a slightly thicker surface. While unexpected at first, this result can be interpreted physically by recalling the underlying modelling of the problem: the optimization provides the best solution given the constraints modelled in the $A$ matrix. Setting the beam pattern to a Bessel function imposes a directivity constraint, making the residual error after optimization more likely to be part of an outlier. Conversely, assuming a flat beam pattern amounts to consider the sensing model as uniform along the vertical axis (no directivity), making the residual error more likely to be spread uniformly along the vertical axis. As a result and due to the natural sparsity of NNLS solutions (Slawski et al. [2013]), this residual error appears grouped with the reconstructed function leading to a thicker surface in this example.

While ringing artefacts outside the objects can be removed using our deringing method, artefacts inside of the object remain a-priori hard to treat in a post-reconstruction processing step. In light of this, choice can be made to adopt a flat beam pattern to minimize deconvolution ringing artefacts as well as reducing the computational cost. On the other hand, a deconvolution with a quantized Bessel function as a beam pattern can be adopted to provide a better reconstruction of the surface at the cost of stronger artefacts.

### 5.6.7 Multiple pass deconvolution

The results presented in the previous section were deconvolution results obtained using a simplified convolution model ($A^{\star}_{ij} = \tilde{\beta}^{i}_{j}$) assuming constant reflectivity along the direction of convolution. Based on this simplified, first-pass deconvolution, the geometry of the object can be identified as well as additional noise and deconvolution artefacts. We investigated here the possibility of using these point clouds as an input to a second deconvolution pass to either provide an estimate of the object normals or information of occlusion.

#### 5.6.7.1 Normal-based kernel

In order to describe better the insonification of the surface of the objects, the reflectivity can be estimated assuming a Lambertian model (see section 2.2.4) and observing the surface normals at each point reconstructed during the first deconvolution pass. As illustrated in figure 5.26, in this situation, the coefficients of the simplified convolution matrix $(A^{\star}_{ij})_{\substack{1\leq i\leq N\\1\leq j\leq P}}$ are scaled by the Lambertian coefficient $\cos(\alpha_j)$ depending on the point of view $\vec{v}(\theta_{ij})$ and the normal at the point considered $\vec{n}$. Following the estimation of

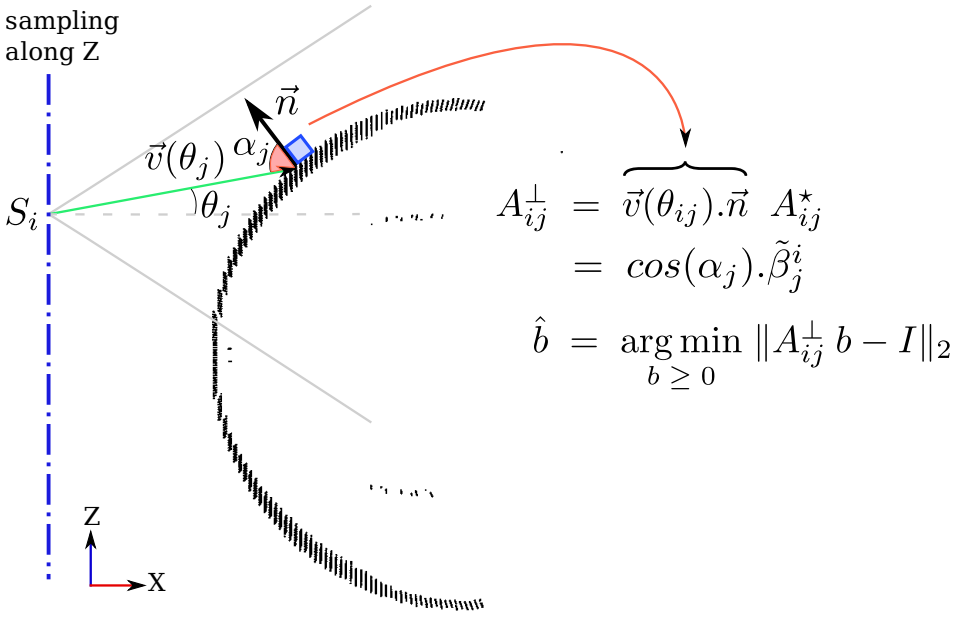



FIGURE 5.26: Multipass deconvolution - normal based deconvolution. When performing a second-pass deconvolution, the normals ($\vec{n}$) at each point can be estimated and used to describe the reflectivity (Lambertian model $\cos(\alpha_j)$) in the convolution kernel.

the normals at each point using an Single Value Decomposition (SVD) description within

a small neighbourhood (5cm), a more refined convolution matrix $(A^{\perp}_{ij})_{\substack{1\leq i\leq N\\1\leq j\leq P}}$ is then obtained by the product of the reflectivity coefficients and the beam pattern coefficients $\tilde{\beta}^i_j$.

#### 5.6.7.2 Occlusion-based kernel

Since the geometry of the object to be reconstructed is *a-priori* unknown, the occlusions occurring during the acquisition of each sample cannot be modelled in the convolution formulation. Once the first deconvolution reconstruction obtained, the geometry of the object is however clearly visible and exhibits many potential occlusions. Similarly to the occlusion resolution algorithm presented in section 4.4.5, this initial 3D representation can therefore be used to provide information on the occluded points to reduce their estimate contribution to the measurement.

As illustrated in figure 5.27, the contribution in the $(A^{\star}_{ij})_{\substack{1\leq i\leq N\\1\leq j\leq P}}$ is modulated by a binary coefficient $\delta_{ij}$. In case of occlusion along the path (source position $S_i$ to the considered

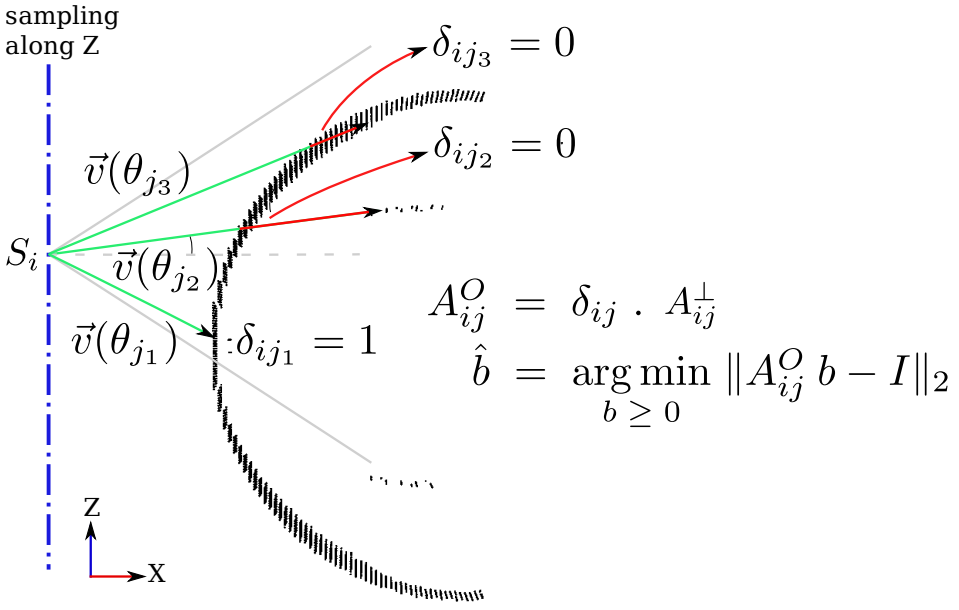
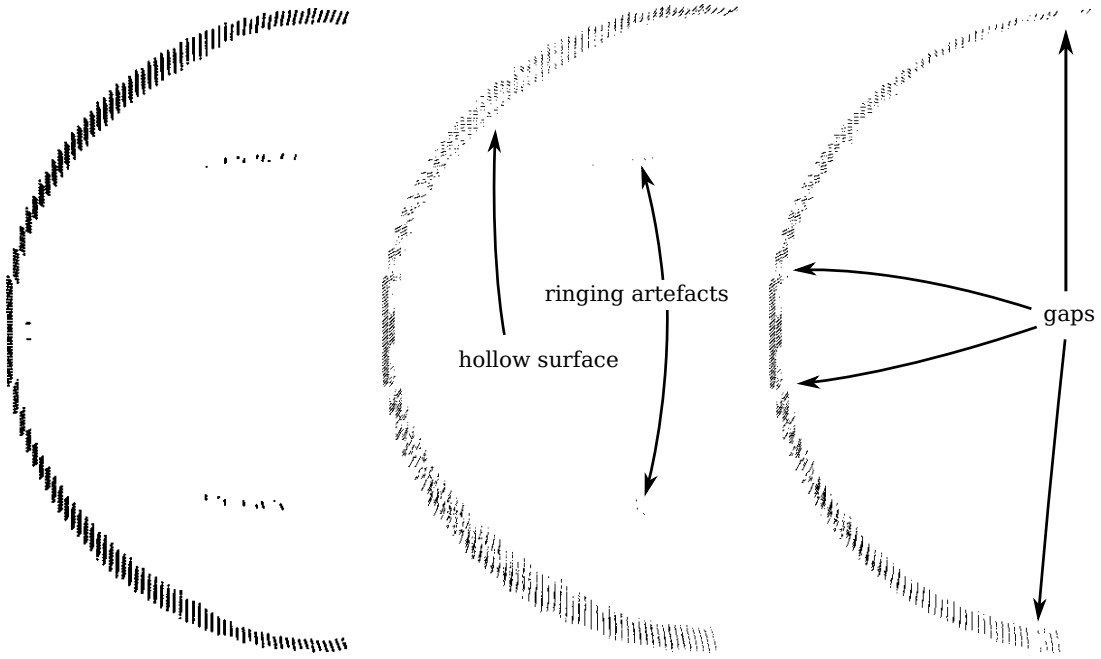



FIGURE 5.27: Multipass deconvolution - occlusion based deconvolution. When performing a second-pass deconvolution, the occlusion occurring along the path of each measurement $\vec{v}$ can be estimated and used to modulate the reflectivity by adding a binary coefficient $\delta_{ij}$ in the convolution kernel.

point), the binary coefficient is set to 0, suppressing the contribution of the point to the measurement. Conversely, in absence of occlusion, the coefficient is neutral ($\delta_{ij} = 1$), enabling a normal contribution following $(A^{\star}_{ij})$ model.

As illustrated in figure 5.28-b, the reconstruction of the sphere using a normal-based kernel enables to reduce the presence of ringing artefacts and makes the surface slightly thinner but still feature a relatively thick and hollow surface which remains unrealistic. Adding the occlusion information (see figure 5.28-c), the surface appears much thinner




(a) Flat kernel (b) Normals (c) Normals + occlusions

FIGURE 5.28: Normal and occlusion-based deconvolution. In comparison to the deconvolution using a flat kernel (a), the addition of normal values in the kernel (b) leads to thinner reconstructions with a few remaining artefacts. Adding occlusion information in the kernel (c) provides a very thin surface but a few gaps appear in the reconstruction.

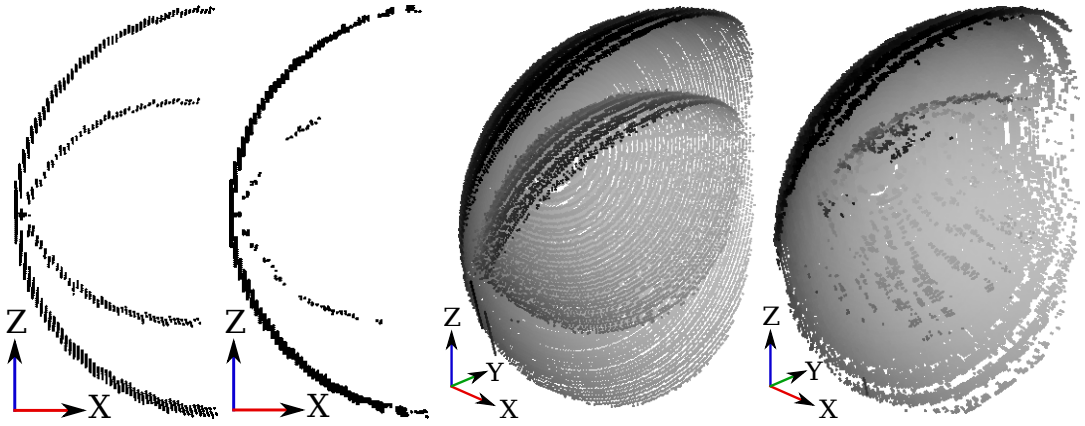
but leads to significant gaps in the reconstructed surface. Although the use of a kernel modelling occlusions and surface normals appears beneficial to the recovery of a thin surface, undesired gaps appear in the reconstruction.

#### 5.6.7.3 Post-reconstruction occlusion resolution

While the deringing method enables the removal of outliers in case where open water was observed, outliers consistently appear inside the reconstructed objects. In this situation, applying the occlusion resolution method described in section 4.4.5 directly on the final point cloud enables to retain the front surface of the object.

As visible in figure 5.29, applying the deconvolution resolution algorithm on the reconstruction enables to remove outliers lying in the inner part of the object and recover a

thinner front surface of the object. On the other hand, the occlusion resolution algorithm




(a) initial reconstruction (b) occlusion resolution (c) initial reconstruction (d) occlusion resolution

FIGURE 5.29: Occlusion resolution on deconvolution reconstruction. In comparison to the initial deconvolution reconstruction (a,c), applying the occlusion resolution on the point cloud enables to remove most of the outliers inside the object (b,c), obtain a thinner surface at the cost of a few gaps on the surface of the object.

remain imperfect: raytracing in the 3D space is done at a limited resolution along both vertical and horizontal aperture, the initial 3D reconstruction is obtained at a limited resolution, leaving small gaps in the 3D representation. Due to these imperfections, a few outliers remain present behind the front surface of the object. Furthermore, gaps inevitably appear in the regions observed at grazing angle such as the top and bottom of the sphere where the vertical resolution hampers the occlusion resolution.

#### 5.6.7.4 Limitations

While these iterative reconstruction approaches present some theoretical interest, their practical use is limited by several drawbacks.

Although simple on noise-free simulated data, the normal estimation, typically based on an SVD estimation, is made complicated on real data due to the presence of noise and outliers. Similarly and in addition to being limited by the raytracing resolution and the first 3D representation resolution, the occlusion estimation is sensitive to noise and outliers. Since it is *a-priori* impossible to completely separate noise from data, the occlusion estimation remains inevitably imperfect on noisy data. These limitations on estimating respectively the normals and the occlusions both result in adding noise to

the convolution model. As a consequence, the deconvolution through the optimization process is made more difficult, requiring increased regularization.

### 5.6.8 Computing resource usage

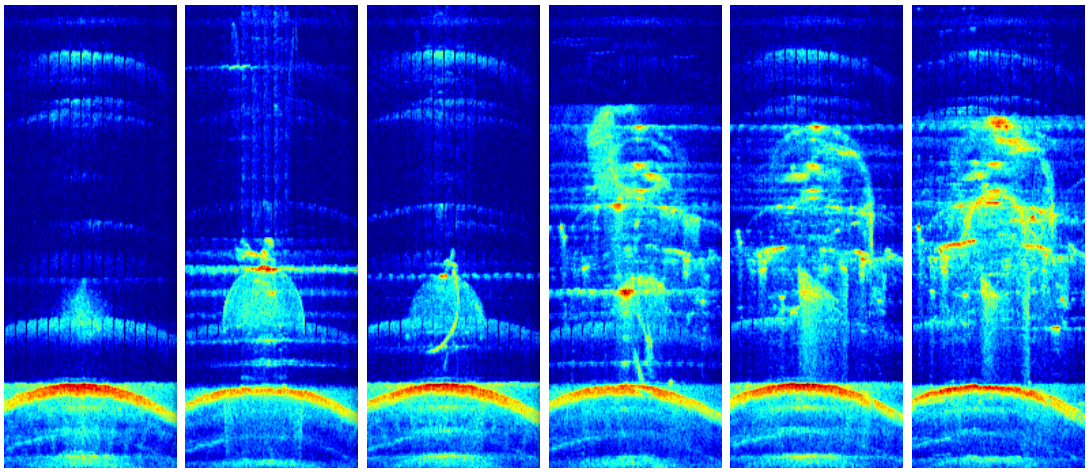
A sequential implementation of our algorithm typically takes about 15 seconds for a real dataset consisting of 128 frames on an Intel Core i7-4700MQ CPU(2.40GHz) processor with 16GB RAM. In comparison, carving takes about 32 seconds for the same data. Both denoising and deringing are important steps in our pipeline (see figure 5.24) but they only require about 1 second of computation each. The least squares sparse solver at the crux of our deconvolution formulation is simple and efficient. Our method is parallelizable, as is the space carving, since we solve independent systems at each $(r, \phi)$ along the $U$-axis.

## 5.7 Analysis

### 5.7.1 3D reconstruction from SONAR for real world observations

As described in the experimental results section (5.6), similar observations were made about simulated reconstructions and real reconstructions. In both cases, the use of a reduced beam aperture or a higher sampling period enabled higher resolutions. Similarly, the use of wide-beam sensors provided increased surface coverage of the objects at every sampling rate and enabled to obtain 3D reconstruction comparable in quality to pencil-beam reconstructions.

**Water tank reconstructions** While these similarities have been observed, additional artefacts appeared visible in the water tank reconstructions. As described in section 2.2, real SONAR imagery typically features various acoustic phenomena such as multipath propagation, crosstalk between different transducers on the same array, transducer induced distortion, environment-related and sensor-specific noise pattern. The dataset acquired in Heriot Watt water tank illustrates these phenomena as showed in figure 5.30. These acoustic artefacts severely pollute the initial 2D images and therefore inevitably impact the quality of the 3D reconstructions. Although impossible to remove using image processing techniques (see section 2.2.6), these artefacts mainly appear due to the

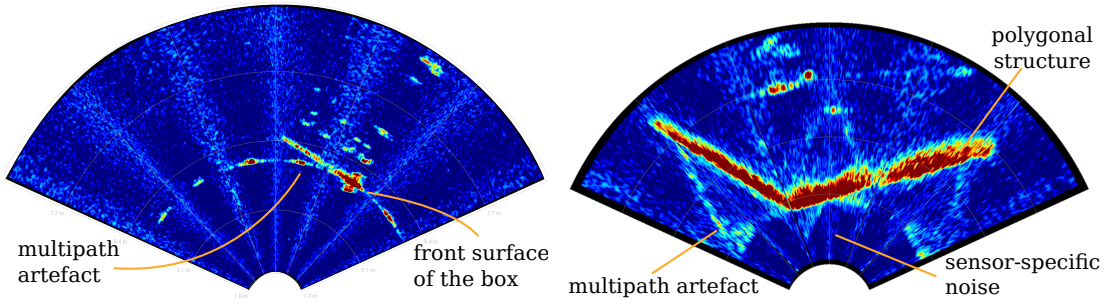


(a) (b) (c) (d) (e) (f)

FIGURE 5.30: Polar images obtained using an ARIS Explorer 3000 sonar in our water tank: a-b-c) Sphere d-e-f) Hyball ROV. The main curve (red and yellow) at the bottom is the ground plane, the inspected objects are observed surrounded by strong multipath artefacts due to operating in a small and closed environment.

presence of a large number of scatterers in the vicinity of the sensor and are therefore environment-dependent.

**Real field observation** When operating on a real field, the environment naturally features less scatterers, reducing the probability of multi-bounce measurements. In this situation, artefacts can still appear but are mainly due to the geometry of the observed structure as illustrated in figure 5.31. Since these artefacts are generated by the structure




(a) box structure inspection (b) polygonal structure inspection

FIGURE 5.31: Images acquired using a Blueview P900-130 sonar during our field experiments: a) Box structure. b) Polygonal structure. Although, in both cases the images feature multipath artefacts, they are created by particular geometries of the inspected structure rather than the environment itself and therefore appear less consistently.

itself rather than the environment (as in the water tank experiment), they appear less consistently and therefore have less impact on the 3D reconstruction process. Although in the case of a sensor embedded on an AUV, the quality of the 3D reconstruction naturally

depends on the accuracy of the navigation data, our experiments showed the ability of obtaining a centimetre-level 3D representation of field structures using standard imaging SONAR and embedded navigation systems. This representation provides enough details for visual recognition of the objects and in some cases structure part identification. As a result, both of our algorithms appear suitable for real world 3D sensing.

### 5.7.2 Deconvolution versus space carving technique

**Performance** As shown on our simulated results, the deconvolution method roughly offers the same reconstruction accuracy as the space carving while yielding better surface coverage with both sensors. Real data experiments confirmed these results and exhibited better regularity of the deconvolution reconstruction. In addition to these differences, the deconvolution formulation provides the possibility of adding regularization in order to manually enforce constraints on the solution such as sparsity along the direction of deconvolution.

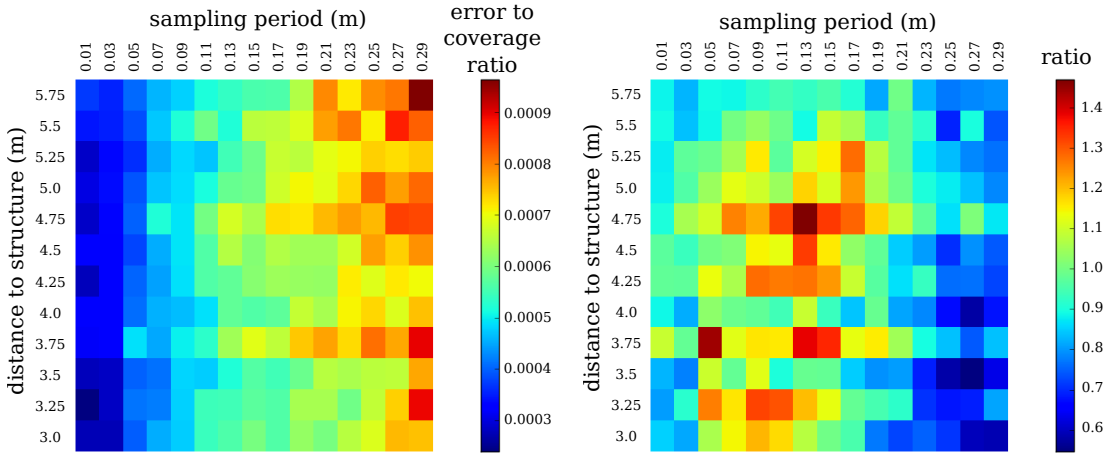
**Practical usage** Along with these qualitative differences, the two algorithms differ by their practicalities: the space carving algorithm enables online iterative mapping by maintaining a temporary 3D representation while the deconvolution algorithm requires to have all the samples available before starting (batch processing). The space carving operates 3D reconstruction from samples acquired along any trajectory and sampling scheme while the deconvolution formulation requires a rectilinear motion and a uniform sampling along the direction of uncertainty of the sensor (vertical aperture).

**Resource usage** Due to their differences in implementation, the two algorithms feature different resources usage. Although based on an efficient storage structure (Octree), the space carving algorithm requires large amounts of memory (up to 16GB) while the deconvolution implementation typically only requires 2GB but appears less scalable (relatively to the number of input images). The computation time of both methods is affected by the presence of noise in the input data: the denoising step on the input samples enables to speed up the reconstruction process. Our two algorithms therefore feature complementary characteristics in term of resource usage.

### 5.7.3 Optimization of the sensing strategy

Both simulated and real data experimental results provided information of interest for the optimization of the 3D sensing strategy.

**A pre-diving tool for sensor selection and parameters optimization** Through the use of CAD models and physics modelling, the simulation environment enables to recreate various real world situations. Thanks to this, both algorithms have been tested in various configurations. In this situation, the use of metrics to evaluate both the geometrical accuracy and the usage of computing resources provide a pre-dive estimation of the performances of the system. Given a CAD model (or a set of CAD models) of the object of interest, the simulation environment enables to explore various inspection configurations. As illustrated in figure 5.32-a, inspection characteristics such as the sampling period, the stand-off distance to the structure or the inspection path can be jointly evaluated. Similarly, sensors of different specifications (vertical aperture, range




(a) Error to coverage ratio of a 10° sensor

(b) Error to coverage ratio of a 10° sensor over a 1° sensor

FIGURE 5.32: Illustration of parameters exploration in a simulation environment on a given situation (inspection of the riser structure). a) Error to coverage ratio obtained with a 10° aperture sensor at various sampling periods and stand-off distances. b) Comparison of the reconstruction accuracy obtained with a wide-beam sensor (10°) and a pencil-beam sensor (1°) exhibiting up to 40% difference.

and bearing resolution, SNR) can be compared as shown in figure 5.32-b. Based on this pre-dive study, a sensor can be chosen for optimizing a given metric (coverage, reconstruction error, inspection time, outliers, error to coverage ratio, error $\times$ inspection time, size of generated data, etc.) while satisfying some hard constraints (minimum stand-off distance, limited degree of freedom while moving the vehicle, etc).

**Wide-beam versus pencil-beam sensor** As presented in the previous section, the choice of the sensor can be explored in specific situations (particular structure to be inspected, constraints on the inspection, etc.) using the simulation environment. In addition to this tool, experiments carried with real data highlighted the differences in 3D reconstruction performances (see section 5.6.1.1). While pencil-beam sensors enable better reconstruction accuracy, the resulting surface coverage remains lower or requires to operate at very high sampling rate. Conversely, wide-beam sensors provide information on large amounts of water, yielding better surface coverage at the cost of a slightly higher reconstruction error.

Overall, the reconstruction results obtained with the two sensors appear comparable and provide elements to choose one sensor or another depending on the application. In situations where priority is given to the accuracy of the results (such as 3D mapping for object recognition), a pencil-beam sensor will be selected. On the contrary, when a slightly less accurate representation can be tolerated (pipeline tracking, occupancy mapping) or when higher coverage of the scene is preferred (limited sampling rate due to a minimum vehicle speed), a wide-beam sensor will be employed.

In most practical situations, the cost and bulk of the equipment are of first concern: the payload of underwater vehicles is often limited by the amount of space available on board, therefore requiring a selection of versatile sensors. In this context, we showed that performing 3D reconstruction from a wide-beam sensor was made possible using our two algorithms with results comparable to pencil-beam. Since wide-beam sensors are commonly embedded on vehicles for performing robust online obstacle avoidance, our algorithms avoid the addition of another expensive and bulky sensor.

Finally and driven by the increasing density of electronic equipment integrated on the vehicles, power consumption is nowadays becoming increasingly more and more important. In this context, a case by case study of each sensor can be performed and compared to their reconstruction performances to achieve the desired trade-off.

## 5.8 Conclusions

In this chapter, we presented a second 3D reconstruction method from imaging SONAR data based on a sparse linear system optimization. Following a formulation of the imaging problem as a convolution along the direction of uncertainty, the 3D reconstruction problem is formulated as a spatially-varying blind deconvolution. In order to address the complexity of the 3D reconstruction problem with a convolution kernel based on the geometry of the observed scene, simple assumptions leaded to a formulation as a constrained optimization of a linear system.

In order to evaluate the performance of the deconvolution algorithm, a quantitative analysis of deconvolution reconstructions from noise-free simulated SONAR data was made, exhibiting similar reconstruction accuracy to the space carving method with increased surface coverage. While featuring visible acoustic artefacts, water tank experiments showed the ability to obtain 3D reconstructions with an increased level of detail in comparison to space carving method. Field experiments demonstrated the ability to use the deconvolution reconstruction method on data acquired by an AUV with limited motion control and navigation accuracy. The influence of regularization was studied, exhibiting increased sparsity as well as limitations in reconstructing flat surfaces along the direction of uncertainty.

A comparative analysis between our two algorithms showed the interest in using the deconvolution method over the space carving method to obtain a better surface coverage. Overall, the two algorithms exhibited comparable performances and featured complementary practical usage constraints. Since maintaining a temporary occupancy map of the scene, the space carving algorithms enables online reconstruction by iteratively adding new information at the cost of a high memory usage. Conversely, the deconvolution method does not require a lot of memory but needs all the input samples at the same time. The space carving method enables reconstruction along any trajectory and sampling scheme while the deconvolution technique requires rectilinear motions and uniform sampling.

We then presented the simulation environment as a pre-dive tool to optimize the sensing strategy and in particular to give elements towards the choice of the most adapted sensor in a given situation. We discussed the pros and cons of using a pencil-beam versus

a wide-beam sensor for 3D reconstruction purposes depending on the priority given on the accuracy of the reconstruction, the cost and bulk of the sensor or the power consumption. In particular, a pencil-beam sensor will offer higher accuracy but requires high sampling rates to achieve good surface coverage and is in general only used for 3D mapping. Wide-beam sensors are on the other hand commonly embedded on underwater vehicles in forward looking configurations for obstacle avoidance and target tracking. In this situation, our algorithms enable 3D reconstruction from wide-beam SONAR data, avoiding the need for additional sensor.

Finally, the 3D reconstruction results obtained with both of our methods exhibit enough detail for visual recognition of the objects or parts of large structures. In particular, our 3D reconstruction offer enough information for obtaining higher level knowledge on the scene.

# Chapter 6

# Object recognition in underwater scenes using SONAR data

> "Para que as coisas existam duas condições são necessárias, que homem as veja e homem lhes ponha nome."
>
> *A Jangada De Pedra* - José Saramago, 1986

## 6.1 Introduction

In chapters 4 and 5, we introduced two 3D reconstruction methods from 2D SONAR images and showed their interest for object recognition and semantic interpretation. Based on these results, we present here a few applications.

We first present a 3D registration technique based on a bundle adjustment of 2D elevation maps of overlapping swathes. Using field data, we show the interest of this technique to compensate for navigation drift when covering large areas, effectively extending our reconstruction methods to the mapping of large underwater areas.

Following the reconstruction of large scenes, we investigate two model-based object recognition methods. The first approach, based on the full 3D point cloud representation, takes advantage of standard matching methods while the second approach proposes

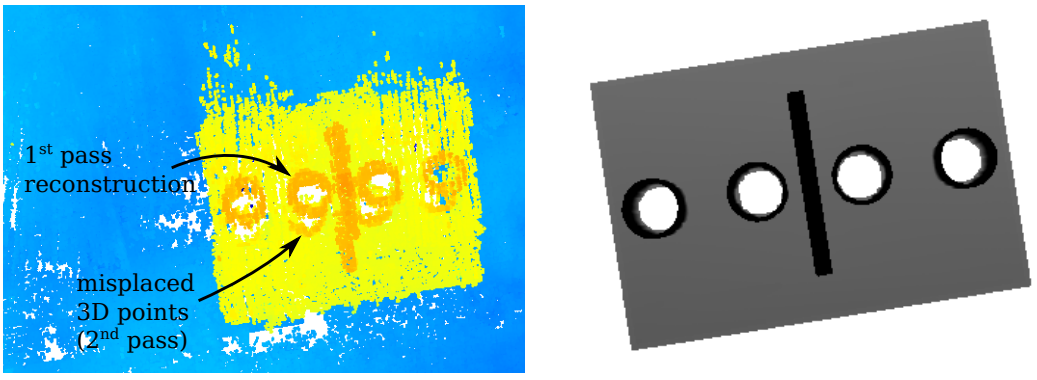
to reduce the complexity of the recognition problem by adopting an elevation map representation. This second method takes advantage of a lightweight rotation-invariant histogram description, offering fast model recognition and matching in large scenes. Using field experimental data, we demonstrate the first CAD-model-based underwater object recognition method.

Once the ability of mapping and labelling large scenes demonstrated, we present a few applications taking advantage of semantic labels. We first demonstrate the generation of lightweight yet accurate maps of the environment. We then discuss applications to the navigation and path planning problems as well as higher-level reasoning for the operation of AUVs in partially structured environments.

We finally explore the possibilities of multi-modal sensing with the mapping of video data on the 3D representation of the environment and discuss some applications.

## 6.2 Large field 3D reconstruction through local registration

Due to significant navigation drift, the 3D reconstructions of the environment from data gathered by an underwater vehicle over distances greater than a few metres exhibit significant artefacts. As illustrated in figure 6.1, in absence of navigation correction between two overlapping 30 metres long swathes, the 3D reconstruction of the structures feature visible geometrical inconsistencies. In these situations, there is therefore a need




(a) (b)

FIGURE 6.1: 3D reconstruction artefacts due to navigation drift. In comparison to the CAD model (b), the 3D reconstruction obtained by superposing two successive swathes (a) exhibit visible artefacts.

to simultaneously perform 3D reconstruction and navigation correction. Since in general, the 3D reconstruction exhibits enough features to allow for manual matching, we propose here a method for automated reconstruction and relocalisation (navigation correction), similar to a Simultaneous Localization And Mapping (SLAM) technique. Similarly to chapters 4 and 5, we assume a local navigation accuracy equivalent to 1% drift, therefore enabling centimeter-level drift accuracy over a meter of travel distance. Due to the necessity of correcting for navigation drift after a few meters, we assume significant overlap is present in the data to allow for manual registration by a human operator.

### 6.2.1 Elevation map generation

In order to reduce the dimensionality of the 6D matching problem, we adopt a 2D representation defined by elevation maps. When observing areas where sensor motion is restricted to a half-space such as when inspecting a seabed, elevation maps offer a lightweight representation of the environment by storing only the highest observed height at each (*North,East*) point, resulting in a (*North,East,Depth*) 2D image. These elevation maps are generated at a given resolution in *North* and *East* and can store depths values in either a floating point representation or at a lower resolution, quantized representation such as unsigned 16bits, reducing the weight of the representation. As illustrated in figure 6.2, a 3D representation requiring 7MB storage can be converted to a 780 kB 2D elevation map representation by subsampling at 2cm resolution along *North* and *East* and quantizing on 16bits in *Depth*. In addition to enabling easier access and processing of the data, adopting a fixed resolution 2D structure representation enables to trade-off accuracy for weight.

### 6.2.2 3D registration through bundle adjustment

In order to correct for navigation drift, we adopt a bundle-adjustment strategy by considering the navigation drift negligible over a few meters but requiring position adjustment after larger scale motion. In the context of a vehicle following a typical lawnmower pattern (see figure 6.3), we propose to relocate each new reconstructed swath relatively to the previous one by operating a 3D matching along the 3 reference directions (*North,East,Depth*). This approach is motivated by the observation that *Yaw* angle estimation is generally very accurate (fraction of a degree accuracy) when using

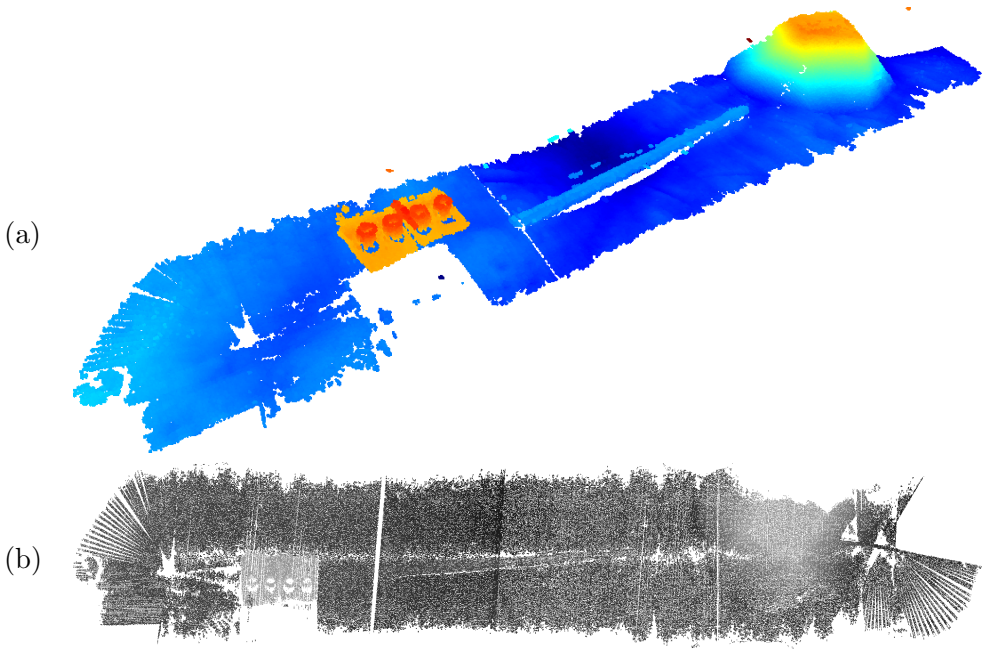



FIGURE 6.2: Elevation map generation. The 37 × 6 metres full 3D representation (a) is converted into an equivalent 2D elevation map generated at 2cm resolution providing a much lighter representation (780 kB vs 7MB).

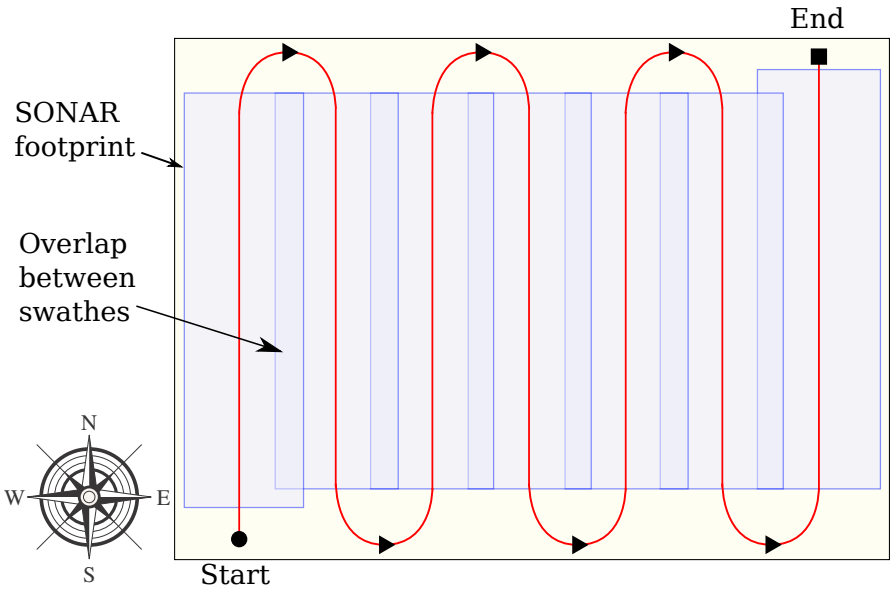



FIGURE 6.3: Lawnmower inspection pattern.

embedded compass systems, the angular variation on the two remaining axis (*Pitch* and *Roll*) remaining negligible.

#### 6.2.2.1 2D matching

The registration process is performed relatively to the origin of the first swath (initial position of the vehicle). We then consider small swathes (30 metres long) which allows us to assume the navigation estimation over this length to be fairly accurate. A registration is then only needed between consecutive swathes when the vehicles operates a U-turn at the end of each line where consequent drift has been observed. In this context, each new swath provides a new elevation map and the registration problem is then formulated as a minimization of the unsigned median distance between two elevation (or depth) distributions. In order to ensure completeness, the distributions are registered by considering only the points appearing in both elevation maps for each given translation $(\delta_N, \delta_E)$ over a finite set of translations corresponding to a discretization over a predefined *North,East* interval $[-\Delta N_{max}, \Delta N_{max}] \times [-\Delta E_{max}, \Delta E_{max}]$ at the resolution of the elevation maps. Due to the low dimensionality of the problem, the position of each new map is tested exhaustively in $[-\Delta N_{max}, \Delta N_{max}] \times [-\Delta E_{max}, \Delta E_{max}]$. Importantly and in order to account for potential elevation offsets due to SONAR miscalibration and pressure sensor inaccuracy, the depth distributions of the overlapping submaps are previously aligned by matching their signed median values (see figure 6.4), providing an offset value $\zeta_{(\delta_N,\delta_E)}$:

$$\zeta_{(\delta_N,\delta_E)} = Median\Big[Z_2(i,j) - Z_1(i+\delta_N, j+\delta_E)\Big]_{\substack{\max(0,-\delta_N)\leq i\leq \min(N_2,N_1-\delta_N)\\ \max(0,-\delta_E)\leq j\leq \min(M_2,M_1-\delta_E)}} \tag{6.1}$$

where $Z_1$ and $Z_2$ are the two elevation maps, respectively of sizes $N_1 \times M_1$ and $N_2 \times M_2$. As described in equation 6.2, the optimal registration $(\delta_N^\star, \delta_E^\star)$ is then selected as the translation resulting in the minimum median value of the unsigned depth distance distribution computed on the overlapping area:

$$(\delta_N^\star, \delta_E^\star) = \underset{\delta_N,\delta_E}{\arg\min}\left(Median\Big[|Z_2(i,j) - Z_1(i+\delta_N, j+\delta_E) + \zeta_{(\delta_N,\delta_E)}|\Big]_{\substack{\max(0,-\delta_N)\leq i\leq \min(N_2,N_1-\delta_N)\\ \max(0,-\delta_E)\leq j\leq \min(M_2,M_1-\delta_E)}}\right) \tag{6.2}$$

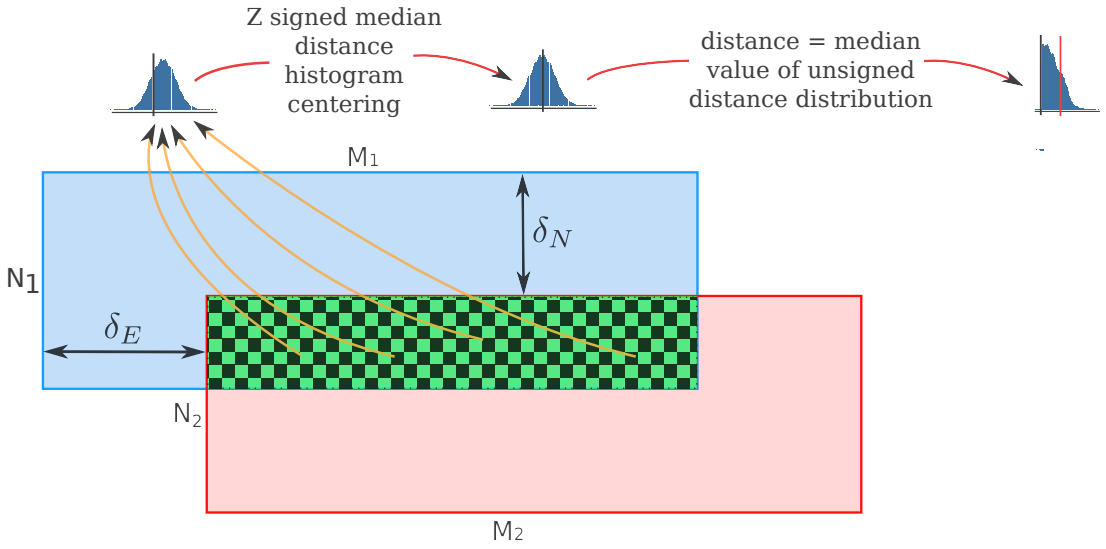



FIGURE 6.4: Median-based swath registration. Following the alignment of depths distribution, the distance between two patches is evaluated using the median of the unsigned distance distribution computed on the overlapping area.

Following this approach, each new map is registered to the previous one and the final 3D map is then given by combining all the co-registered 3D swathes.

#### 6.2.2.2 Experimental results

We present in this section field experimental results based on a dataset acquired in Loch Eil in Scotland. The data was gathered by the Subsea7 AIV (see section 4.7.1 for description) prototype during validation trials. Prior to the experiment, three oil-field type structures as well as two pipelines were laid on the seabed with various orientations. The structures and the pipelines were deployed on the seabed forming a field area of 50 $\times$ 34 metres. As visible in figure 6.5, the structures were laid with arbitrary orientations with pipelines placed in between each structures mimicking connectivity between the structures. The pipelines appeared half buried in the seafloor composed of mud and rocks. As visible in figure 6.6, a lawnmower pattern was followed by the AUV along the East-West axis. In this configuration, the full field was inspected in an hour time. The data used for this experiment was acquired by the BlueView MB2250 SONAR mounted in a downward configuration featuring a pencil-beam vertical aperture of 1° and an horizontal aperture of around 80°. The along-track sampling period was 4cm with an average range to the seabed of 5m. Due to the small overlap between each sample, the space carving reconstruction technique (see chapter 4) was used. The navigation

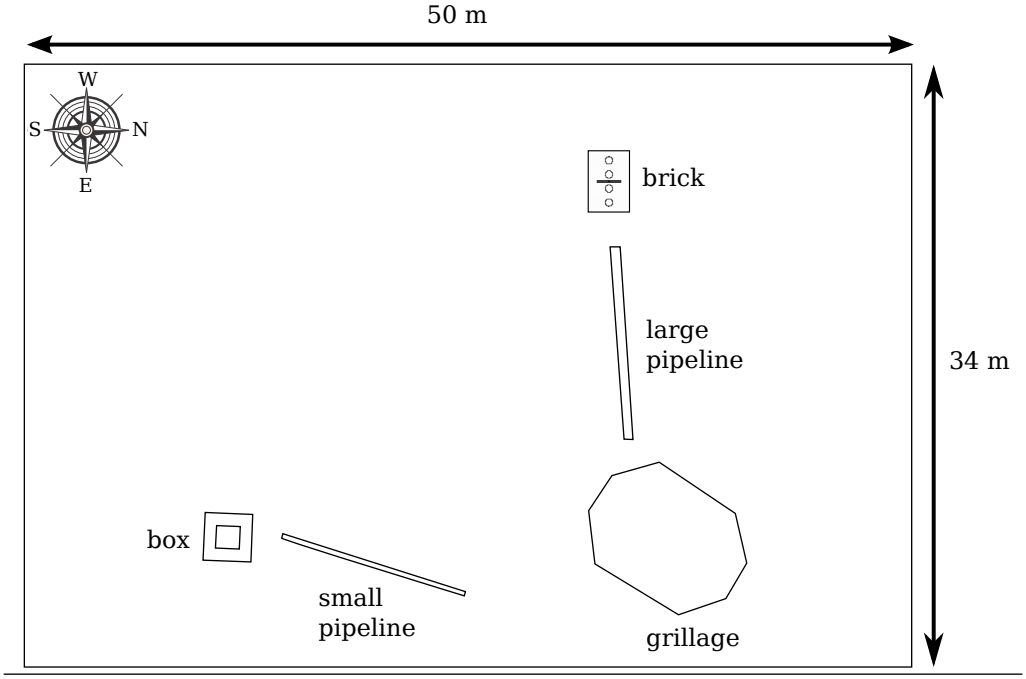



FIGURE 6.5: Loch Eil trials field map exhibiting the three reference structures and two pipelines connecting them.

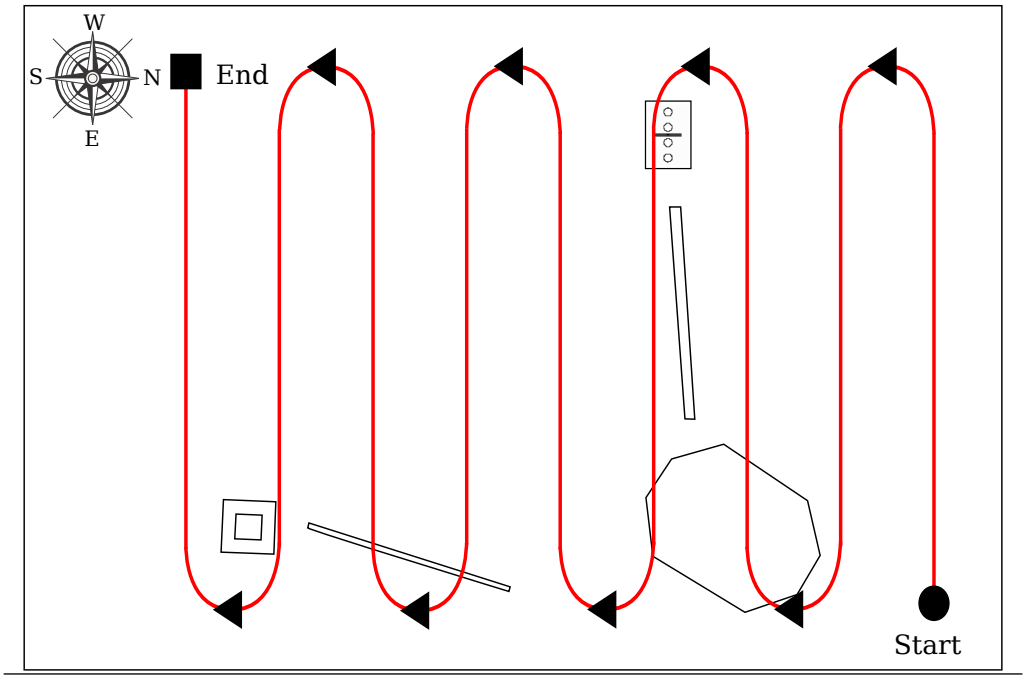



FIGURE 6.6: Loch Eil trials field inspection pattern. The pattern lawnmower pattern followed by the AUV enables a top inspection of the full field in an hour time.

data was provided by a module integrating the readings of three sensors: bottom-lock velocities were provided by a DVL (Doppler Velocity Log), depth measurements were given by a pressure sensor while the estimation of the vehicle orientation relied on a compass and a gyroscope.

The swathes elevation maps were generated at 1cm resolution in *North/East*. Since the along-track sampling period is only 4cm on average, some gaps are apparent on the map. The bundle adjustment was therefore operated at 1cm resolution, discarding undefined points. To illustrate the benefits of the registration between swathes, figure 6.7 shows the 3D reconstructions of two structures and a pipeline before (figure 6.7a-c-e) and after registration (figure 6.7b-d-f). In the case of the original example depicted in figure 6.1, the registration leaded to a shift of 24cm in *North* and 2cm in *East* (figure 6.7-b vs 6.7-a). While assembling all the swathes, the whole field was reconstructed with a maximum registration shift of 38cm in *North* and 18cm in *East*. The full reconstruction of the field is depicted in figure 6.8. In addition to the three structures and two pipelines deployed on the field, the reconstruction exhibits two trenches of about 30cm depth.

**Accuracy assessment**

While the benefits of applying our SLAM technique on this experiment can be visually assessed in figure 6.7, no accurate prior map of the field was given as a comparison. No additional localization system such as USBL was deployed on the field of trials, making a quantitative analysis restricted to the comparison of each structure to its rough CAD model. When doing this, the CAD models need to be aligned to the 3D reconstruction requiring either manual input or an automated matching method. In both cases, the distance obtained by comparing the point cloud to its registered model includes the 3D reconstruction error, the registration error and the model matching error. For this reason, we choose to provide details on these values in section 6.3.2.7.

#### 6.2.2.3 Applications

SLAM-issued 3D reconstruction point clouds provide direct visual rendering of the environment of interest to human operator for exploration, operation planning such as man-made structure or vehicle deployment. In addition to this, the 3D representation can be used as an input to automated online navigation and path planning.

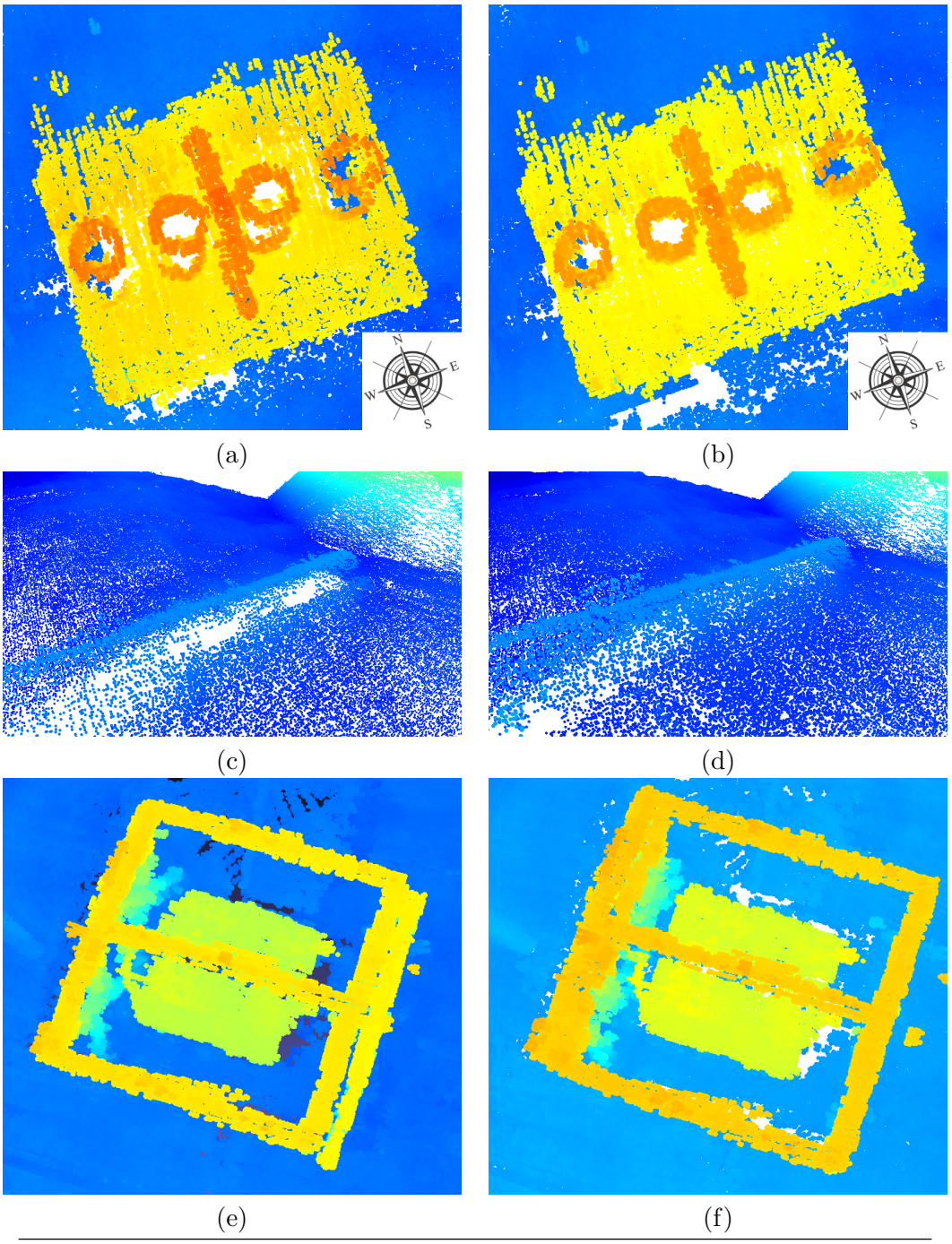



FIGURE 6.7: Illustration of field 3D reconstructions after swathes registration. a-c-e) 3D reconstruction of a structure without registration. b-d-f) Registering the two swathes containing the structures (a,b and e,f) pipeline (c,d) enables to recover a consistent 3D reconstruction.

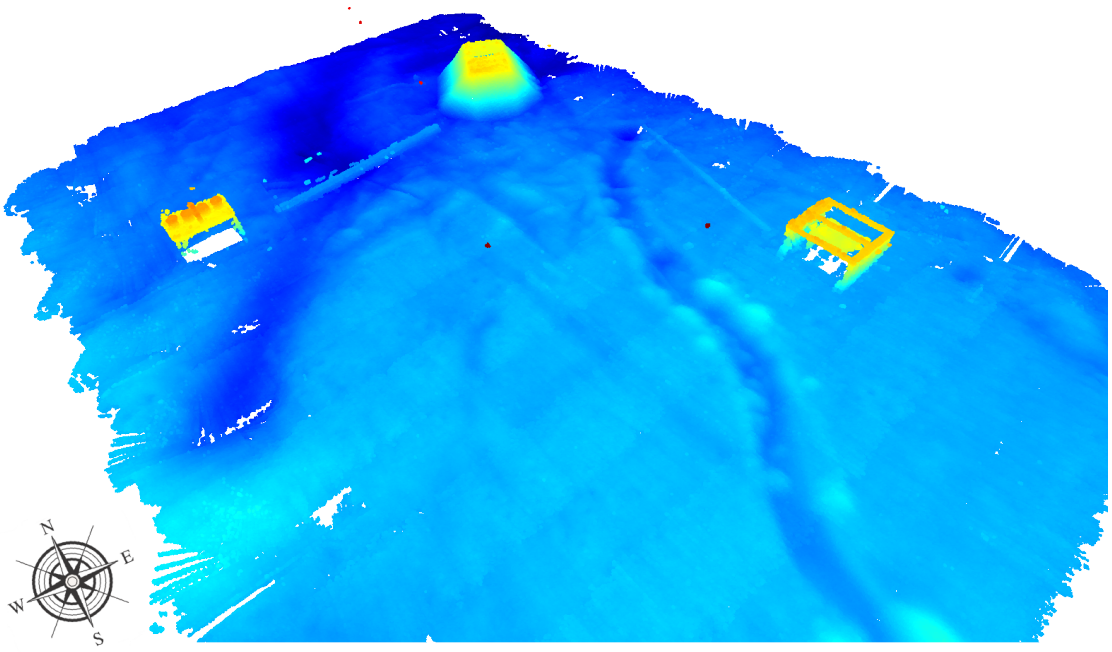



FIGURE 6.8: Complete reconstruction of the field of trials in Fort William.

The first outcome of applying a SLAM method while inspecting a field is the ability to correct for navigation drift using observation. These so-called terrain-based navigation methods have been widely studied (Kinsey et al. [2006]), both in presence of prior knowledge on the environment (Eustice et al. [2005], Williams and Mahon [2003]) or with concurrent generation of landmark maps (Roman [2005]). These methods therefore enable accurate positioning from 3D sensing.

Similarly, 3D space occupancy information is of interest to vehicle motion planning by generating exclusion zones and applying optimal path planning methods such as navigation functions (Latombe [2012]).

The 3D point clouds resulting from reconstruction of the field do not provide any direct semantic information but the structures visible in our field results are clearly identifiable by a human operator making possible a manual labelling of the field by an operator (such as the map presented in figure 6.5) in order to assist higher level tasks. Due to the variability of appearance of the objects in the scene, the potentially large size of the reconstructed area and the cost in time induced by vehicle recovery for data analysis by a human operator, there is a great interest in being able to obtain semantic information on the scene automatically.

## 6.3 CAD-model-based object recognition

Following the extension of our local 3D reconstruction methods to larger scale inspections, we propose to extract semantic knowledge from the 3D representation of the environment. Due to the difficulties of acquiring an accurate 3D representation of underwater environments, most underwater semantic mapping approaches have been based on texture characterisation rather than geometry analysis. Reflectivity maps have been estimated to characterise seabed areas (Chen et al. [2008]) and marine habitat (Kenny et al. [2003]). The estimation of the physical properties of sediments typically require SONAR sensors operating at low frequencies (10Hz to 10 kHz) to allow for subsurface penetration (Schock [2004]). Low-frequency designs limit the range accuracy of the sensor therefore its usability for the estimation of the 3D geometry of the environment.

When operating at higher frequencies, the geometrical information has been used to classify marine vegetation (Mizuno and Asada [2014] using a texture classification in the 2D images. Following the idea of adopting a plane-based representation (Birk et al. [2010]), a two-level semantic annotation of the environment was presented in Pfingsthorn et al. [2011]. In this study, planes were first fitted to operate a rough classification of the areas into generic categories (seabed, wall) while a second step focused on the type of object by analysing the local normals distributions. While interesting, this method is limited to providing a global characterisation of the environment but do not provide fine semantic information inside each area. In particular, this approach does not address the detection of man-made objects in the scene.

In contrast to these studies, we propose to extract semantic knowledge from the 3D representation of underwater environments by operating a 3D matching of a set of rough CAD models representing known man-made structures or parts. We formulate the problem as a 6D problem where the scene can be observed as full 3D such as a vertical pipeline around which the vehicle could turn to see both sides. In this context both translation and rotation information of the structure are unkown but we assume rough knowledge on the size of the pipeline / buoyancy. Note that due to the limited resolution of underwater observations and the potential presence of marine growth, the shape and size of submerged objects varies in time and can therefore only be assumed to be known to a limited level of accuracy (centimer-level or more).

### 6.3.1 Model recognition in the 3D scene

While the most common approach to perform object recognition in the air is feature description and matching (Bay et al. [2006], Drost et al. [2010], Johnson and Hebert [1999], Lowe [1999]), these techniques exhibit several drawbacks prohibiting their use for model-based object recognition in underwater reconstructions. Since feature description methods aim at characterising the local geometry around a point, most of the descriptors require computation of surface normals. The presence of noise and outliers in the 3D reconstruction of underwater scenes makes the estimation of normals challenging if not impossible. In addition to this robustness challenge, the surface of objects submerged for more than a few weeks or months commonly features marine growth or dirt. As illustrated in figure 6.9, marine growth significantly alters the appearance of the structures by modifying their shape. Finally due to the variability of the appearance of objects when

FIGURE 6.9: Marine growth on the buoyancy of a riser structure. The cylindrical shape of the original buoyancy now appears very irregular.

observed with SONARs, the feature description of an underwater measurement needs to be robust to noise. For this reason and although 3D reconstructions from SONAR feature limited accuracy, the use of a higher resolution sensing modality in presence of marine growth would not simplify the object recognition. In particular, the a-priori knowledge on the geometry of the object of interest is typically is at best expected to consistitude in a CAD model of the object before deployment. Therefore due to the simplicity of the initial CAD model and the variability of the appearance of submerged object, the description of the local geometry does not provide any reliable information.

Following unsuccessful initial experiments using standard feature-based methods (Bay et al. [2006], Rublee et al. [2011]) and due to the low resolution and noisiness of underwater reconstructions, we chose to address the object recognition problem by applying direct model matching techniques. Although naive, this approach will constitute a reference to compare to the method described in section 6.3.2. Direct 3D matching techniques are prone to sub-optimality and computationally intensive due to the high dimensionality of the model matching problem: 3 degrees of freedom in rotation, 3 degrees of freedom in translation and one more degree of freedom for the choice of the candidate object. In order to address this, we reduce the number of model matching to operate by dividing the scene in subspaces and discarding the candidate subspaces with unrealistic characteristics.

#### 6.3.1.1 Scene partitioning and subspace analysis

In order to address the complexity of matching in 3D, the search space (full scene) is divided into equally sized subspaces. In order to avoid privileging any direction when sampling, each subspace is spherical with a radius $r_{subspace}$ defined as the minimum size allowing any 3D rotation of the model inside the subspace. $r_{subspace}$ is therefore defined by the size of the object and the 3D space "lost" while sampling the scene at a low resolution (grid spacing between two subspaces):

$$r_{subspace} = r_{model} + \frac{\sqrt{3}}{2} . T \tag{6.3}$$

where $r_{model}$ is the radius of the bounding sphere of the model (distance between its centre and its furthest point) and $T$ is the sampling period, generally expressed as a fraction of $r_{model}$.

Following this sampling scheme of the scene, each subspace represents a region potentially containing the object of interest and at least one of them contains the whole structure. Depending on the size of the object in comparison to the size of the scene, a large number of subspaces can be generated at this stage, making a direct 3D matching on each of them prohibitive. In order to decrease the number of potential candidates, each subspace is first evaluated on simple metrics representing necessary conditions for representing the object.

**Oriented Bounding Box (OBB)**

The OBB of the model must fit in the OBB of the subspace. The two OBBs are computed based on their Singular Value Decomposition and the centroid of the each point cloud. The sizes of the sides are then obtained by measuring the furthest point along each direction. The 3 resulting values are finally sorted by decreasing order and compared between the two point clouds. Storing these values in two 3D vectors $\ell_{subspace}$ and $\ell_{model}$, we therefore respectively have $\ell_{subspace}[1] \geq \ell_{subspace}[2] \geq \ell_{subspace}[3]$ and $\ell_{model}[1] \geq \ell_{model}[2] \geq \ell_{model}[3]$. We then evaluate the following criteria:

$$\begin{aligned} \ell_{model}[1] &\leq \alpha_{fit} \cdot \ell_{subspace}[1] \\ \ell_{model}[2] &\leq \alpha_{fit} \cdot \ell_{subspace}[2] \\ \ell_{model}[3] &\leq \alpha_{fit} \cdot \ell_{subspace}[3] \end{aligned} \tag{6.4}$$

with $\alpha_{fit}$ being a margin coefficient.

**Minimum volume**

Based on the OBBs, the volume of the bounding box of the subspace can be compared to the volume of the model. Due to noise and potential background around the object (seabed, rope, etc.), the volume of the subspace is expected to be high enough compared to the volume of the model:

$$V_{subspaces} \geq \alpha_{volume} \cdot V_{model} \tag{6.5}$$

with $\alpha_{volume} < 1$ a margin coefficient.

**Number of points in the cloud**

A minimum number of points in the cloud is desired to represent the geometry accurately. This criteria is typically designed to discard subspaces containing only a few outliers points. Formally, the decision is made based on a simple threshold:

$$N_{subspaces} \geq N_{min} \tag{6.6}$$

with $N_{subspaces}$ and $N_{min}$ respectively being the number of point in the subspaces and the minimum number of points allowed for representing accurately the object.

**Space occupancy**

By parsing the point cloud in a regular grid representation, an estimation of the volume

of the point cloud can be measured by counting the number of occupied voxels. The two occupancies $O_{subspace}$ and $O_{model}$ can then be compared by:

$$O_{subspace} \geq \alpha_{occupancy} \cdot O_{model} \tag{6.7}$$

with $\alpha_{occupancy}$ the margin coefficient on occupancy, typically chosen as lower than 1. If one of these criterias is not satisfied, the subspace is then discarded as not realistically representing the model. On the contrary, subspaces satisfying all these minimum thresholds are kept for further evaluation using a direct 3D matching method.

#### 6.3.1.2 Direct 3D matching

3D model matching techniques are typically divided into two main categories. RANSAC-inspired, voting techniques where each degree of freedom is quantized, triplets of points are chosen respectively in the model and the data point clouds, the transformation is computed and a vote is added in a voting table. The optimal transform is then obtained selecting the transform with the highest number of votes. Although independent of the initial positions of the model and data point clouds, these methods are computationally expensive. The second approach to the matching problem aims at solving the correspondence problem: given a point in the data point cloud, find a point in the model point cloud that corresponds to it. Once enough correspondences are computed, a rigid transform can then be computed in such a way to minimize the distance between corresponding points. The most common way to associate 3D points relies on the description of the local geometry and feature matching. These methods require accurate 3D representation with low noise levels and high similarity between the model and the scene which cannot be assumed in the case of underwater environments.

In this situation, a simple approach is to consider the closest points in the dataset as correspondences. Following this idea, the ICP (Iterative Closest Point) method (Besl and McKay [1992]) iteratively perform registration between two point cloud by minimizing the $L_2$ distance between each point and its closest neighbours. Since its introduction in 1992, many variants emerged, decreasing the time needed for registration (Rusinkiewicz and Levoy [2001]) or increasing robustness to noise, outliers and sparse datasets (Trucco et al. [1999]). While ICP allows for fast registration between two close point clouds, it is not guaranteed to converge to the global optimal registration but rather the closest

optimum. This technique is therefore restricted to situations where the two points clouds are already relatively well positioned. This is the case for local registration applications such as aligning two successive acquisitions from a sensor and provide some sensor motion correction. In order to address this sub-optimality issue, global optimum ICP methods have been investigated such as the Go-ICP method (Yang et al. [2013]). The Go-ICP algorithm achieves global optimality by using a branch and bound technique to explore the 6 degrees of freedom space composed by $SE(3) \times SO(3)$. This branch and bound technique enables an efficient exploration of the solution space by bounding the registration error and focusing on solution subspaces. The ICP technique is then used to perform a fine local registration. This method proves to outperform ICP in terms of registration error at the cost of a longer processing time. Resilience to outliers as well as the application to 3D object localization have also been investigated by the authors with promising results.

Considering the characteristics of underwater 3D reconstructions, we choose to operate model matching in each subspace using the ICP technique when some prior information is known about the most likely orientation of the model in the scene, with low clutter generating local minimas. On the contrary, the Go-ICP method can be chosen when no initial assumption can be made on the scene. The C++ implementation of the ICP method was provided by the PCL library (Rusu and Cousins [2011]). The registration of CAD models to the 3D points clouds of the scene is operated by first operating a regular sampling of the CAD model at a high resolution (1cm), applying a subsampling at a fixed resolution (typically 3cm) on both points clouds in order to reduce the amount of computation and finally operate a point-cloud-to-point-cloud registration. We modified the C++ implementation of Go-ICP provided by the authors of Yang et al. [2013] to allow for shorter branch and bound search by adding a user-defined threshold on the lower bound, representing the expected matching accuracy.

#### 6.3.1.3 Experiments

We performed experiments on both simulated and real data to evaluate the accuracy and resource consumption of this approach. The first set of experiments was performed on 3D reconstruction issued from noise-free simulation data. The 3D reconstruction point clouds were representing a vertical pipeline (riser) with one or multiple buoyancies for

which a rough CAD model (half a cylinder) was provided. In the case where multiple objects are expected to be present in the scene, a threshold on the error to CAD model after matching is set to represent a decision on whether or not the object of interest was found. The subspaces were generated at the period $T = 0.5.r_{model}$, the OBB fitting criteria was set to $\alpha_{fit} = 1.2$, $\alpha_{volume} = 0.6$, $N_{min} = 10$ and $\alpha_{occupancy} = 0.6$. As illustrated in figure 6.10, this method enables to successfully detect the buoyancies in the point cloud using ICP on respectively 8 (out of 19) and 18 (out of 59) subspaces with an average matching error of 3.5cm. Both detections are achieved in less than a second

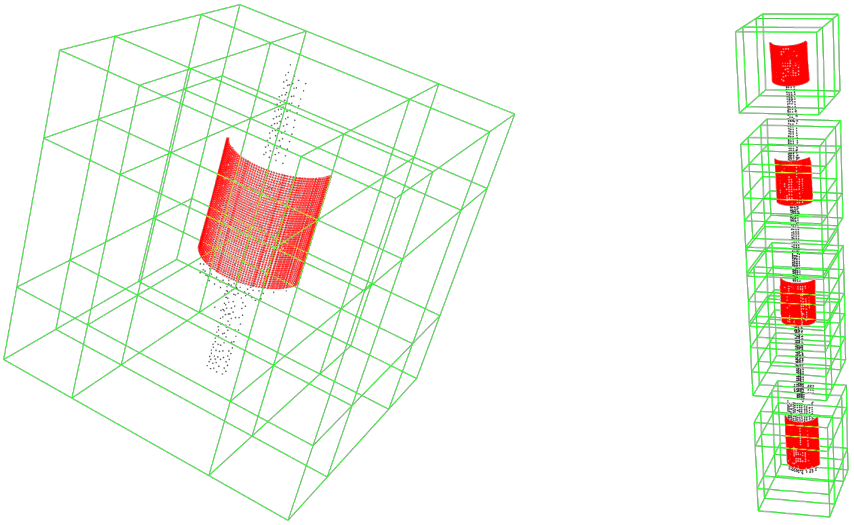



FIGURE 6.10: Buoyancy detection in 3D on simulated data. When inspecting a vertical structure, the ICP method enables to detect the position of one (a) or multiple buoyancies (b) using the subspace analysis and the ICP method. The subspaces are represented by their bounding boxes in green while the matched CAD models are represented in red.

on an Intel Core i7-4700MQ CPU(2.40GHz) processor, typically using under a 100MB memory.

The second set of experiments was performed on field 3D reconstruction issued from AIV data, naturally featuring higher level of noise and outliers. We first applied our ICP-based 3D reconstruction method on two different datasets where the object is clearly visually identifiable. The first 3D reconstruction was presented in section 4.7.3.2 (figure 4.34) and features a pipeline, two mooring chains and a buoyancy at the top of the reconstruction. As visible in figure 6.11-a, the buoyancy is successfully detected. The detection algorithm discarded 144 (out of 226) subspaces and provided the best match of a 50cm diameter half cylinder in 19 seconds with an average matching error of 5.4cm, the model being initially oriented along the vertical direction making a 30° angle with

the orientation of the buoyancy in the point cloud. Similarly and as illustrated in figure 6.11-b, we applied the pipeline detection on a 34m long pipeline reconstruction where the pipeline was standing above the seabed (20cm higher) with a 30cm diameter half cylinder CAD model. In this situation, the pipeline was detected all the way through the point

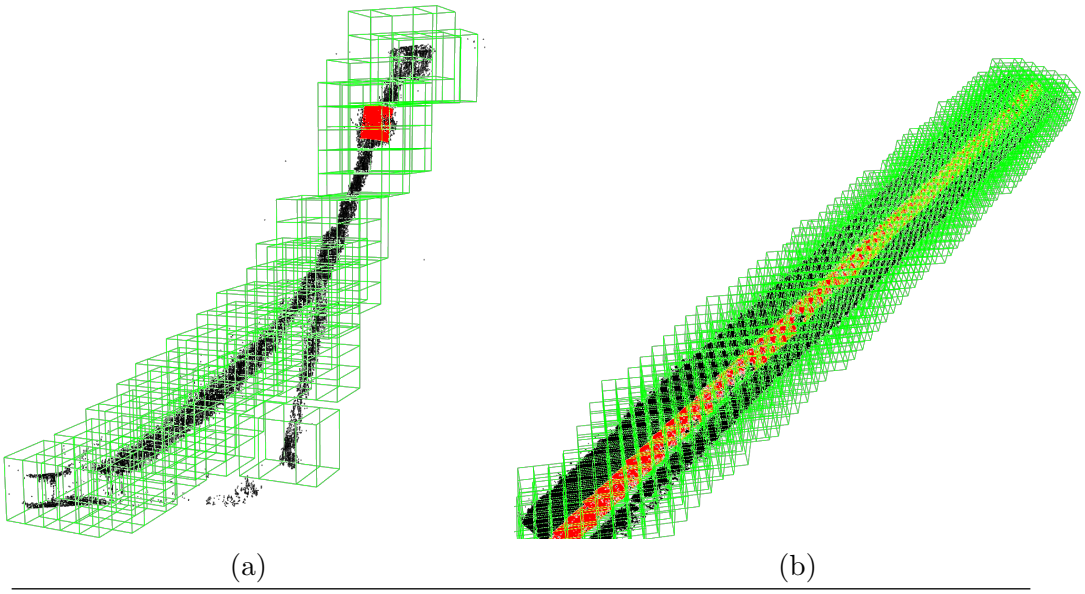



FIGURE 6.11: Buoyancy detection in 3D on field data. The buoyancy present in the riser reconstruction (a) is detected succesfully in a 19 sec while the 34m long section pipeline takes nearly 20mn processing time for recognition in similar conditions.

cloud in 3 minutes and 36 seconds. When the object of interest can be expected to stand out from the rest of the scene (low clutter) such as in the two previous examples, the ICP-based recognition algorithm provides a fast and efficient way to detect the object. In order to evaluate the performance of this technique on a more complex situation, we applied it to a reconstruction of a buried pipeline. In this situation, the pipeline is covered by either sand or a protection (rock mattress) making its recognition significantly more challenging.The typical curve representing a buried object is then surrounded by a seabed which can *a-priori* feature various geometries. The CAD model was initially orientated along the axis of the subspace grid, therefore rotated by 45° relatively to the pipeline in the point cloud. As illustrated in figure 6.12-a, applying the ICP-based recognition algorithm enables to detect the pipeline at each point in the point cloud. The orientation of each instance of matched CAD model appears slightly misaligned (up to 20° ) in some cases due to the influence of the seabed points. In this situation, replacing simple ICP registration by a globally optimal Go-ICP method (see figure 6.12-b) enables a better registration at the cost of a longer processing time (21 mn versus 8 min 23

sc). The use of Go-ICP also leads to outliers in the detection process as visible in figure

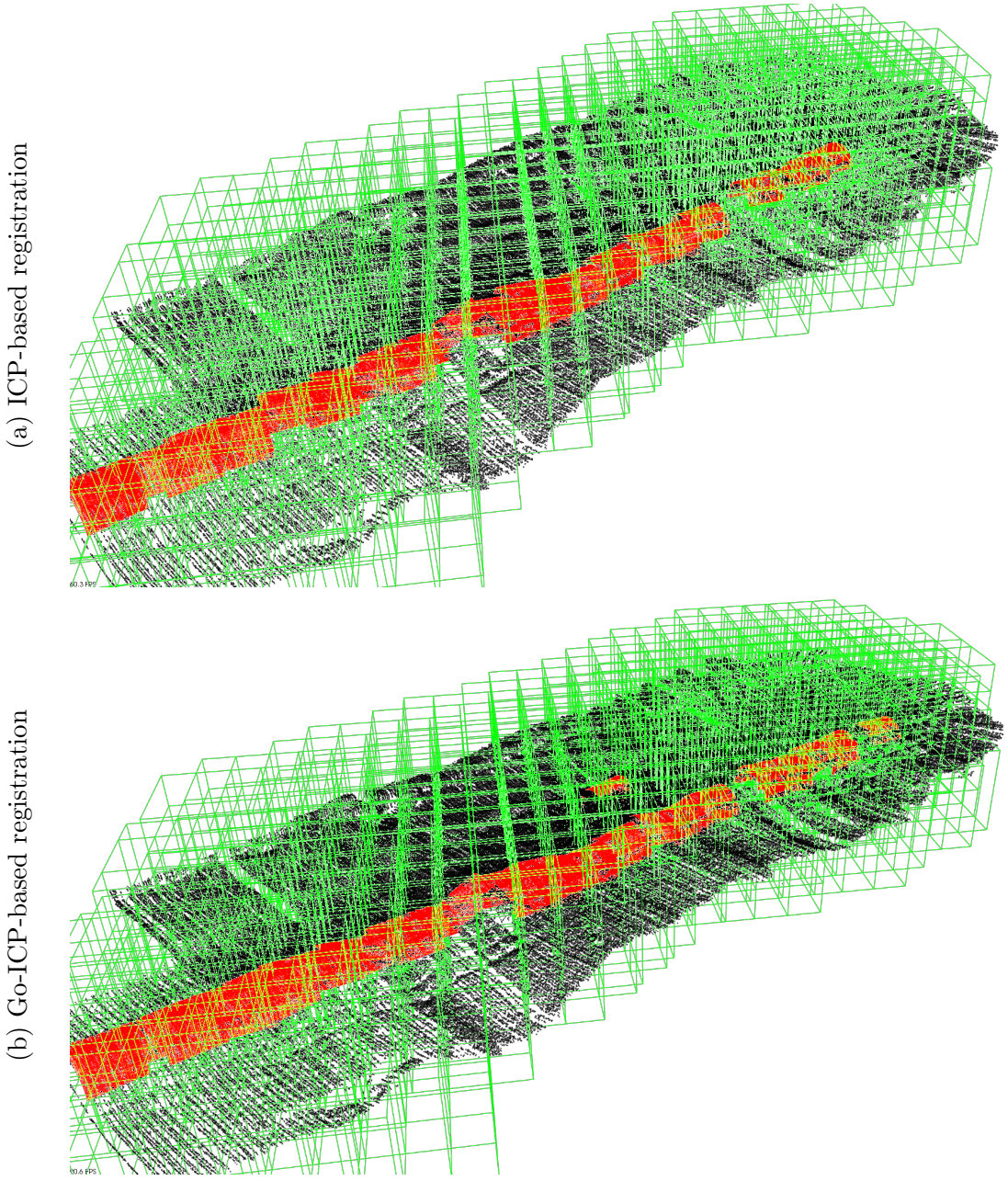



FIGURE 6.12: Buried pipeline detection on field data with both algorithms: ICP (a) and Go-ICP (b). Registration with ICP provides lower angular accuracy due to the influence of seabed points creating local minimums whereas the use of the globally-optimal Go-ICP method provides better registration at the cost of a few outliers due to geometrical similarity with the seabed.

6.12-b where a detection was obtained on the left side of the main pipeline due to the curvature of the seabed in this area.

In order to evaluate the recognition of objects in larger scenes, recognition results of model matching to the field reconstruction of Fort William trials (see section 6.2.2.2,

figure 6.8) are presented in appendix 6.7. In this situation the localization was only partially successful (2 structures out of 3) and required prohibitive computing durations.

#### 6.3.1.4 Applications and limitations

In light of these experiments, direct applications of this simple model matching technique can be deducted as well as some practical limitations.
As shown by the experimental results, it is possible to recognise simple shapes such as pipelines and buoyancies in real field 3D reconstructions obtained from SONAR data. These results show the ability to perform model-based semantic labelling of the 3D scene. The ability to detect pipelines consistently in different configurations allows 3D pipeline tracking. When additional prior knowledge is available, the ability to detect parts of the structures such as buoyancies on the riser dataset could enable online relocalisation of the AUV relatively to these elements as well as part-specific application such as leak detection or manipulation.

In practice the detection accuracy is limited by the sensitivity to local optimums of the ICP method. Although a globally-optimal method such as Go-ICP can address this problem, this comes at the cost of an increase in computation cost. More generally, the scalability appears to be the main limitation of this method as illustrated with the buried pipeline example where a few minutes are required to process an area inspected by the AUV in less than a minute. Although a few minutes of processing time can be acceptable when detecting one structure, a more general application scenario would involve a library of model of potentials objects of interest (with possibly different sizes) making semantic mapping from a large library computationally expansive. In addition to this, while large objects appear to be easily detectable due to their larger size compared to the whole scene (fewer subspaces), they also feature more points making the matching process more costly. On the other hand detecting a small object in a comparatively large field results remains computationally intensive process due to the high number of candidate subspaces. As illustrated in section 6.7, structure recognition on Fort William field was only partially successful and required large computing durations. A common practice to reduce the computational footprint when performing 3D data processing is to operate a prior downsampling of both the model and the data point cloud but this comes at the cost of lower registration accuracy and in particular the possibility to miss

the initial global optimal registration. When considering large number of points such as the 6 millions points of this field, a considerably larger amount of memory is needed (typically 3.5 GB). Experimental results also showed the limited efficiency of the subspace analysis conducted prior the matching due to the sensitivity to the appearance of the input data: when buried, only part of the object might be visible on one hand but the seabed leads to a large bounding box on the other hand. The simple criterias on volume and occupancy are sensitive to the initial coverage of the structure, itself depending on the type of inspection and sensor operated.

### 6.3.2 Structure recognition in 2D space

In order to address the scalability issue, we chose to approach the recognition problem using a lighter representation. Most man-made underwater structures or objects can be recognised in 3D reconstructions from SONAR samples acquired in half a 3D space: seabed inspection is naturally restricted to motions in half a space and typically performed by acquiring range observations along a lawnmower pattern (see description in section 6.2.2.2), single-pass inspections of vertical structures (such as the vertical riser inspection presented in figure 4.34) typically provide range information acquired along a 3D axis, the inspection of ship hulls or vertical walls typically feature similar properties. In this situation, the aspect of the objects of interest can be generally assumed: offshore structures can be laid on the seabed with an arbitrary yaw angle but cannot be deployed upside-down. Therefore the objects can typically be expected to be observed under 4D points of view (3 translations, one rotation (yaw angle) on the plane separating the two half-spaces) by neglecting rotation on the two remaining rotation axis. A full 3D representation such as a point cloud is then unnecessarily detailed and memory consuming. In light of this, we chose to adopt a 2D representation of both the scene and the model by using elevation (range distance) maps as presented in section 6.2.1. In order to address the problem of variable appearance of underwater observations described in section 6.3.1, we chose to adopt a global description of small 2D patches rather than a local (pointwise) feature description.

#### 6.3.2.1 Overview of main feature description methods

Feature description aims at characterising a set of samples by operating a projection from the data domain (images, 3D point clouds, etc.) to a feature domain, represented by a vector of descriptors. The set of samples to describe is typically defined as the neighbourhood of a point by selecting a fixed number of points or all points within a given radius. The description then consists in characterising the neighbourhood through a simple operation applied on the input data. Importantly, the descriptor needs to be specific enough to capture the information of interest (such as the shape of the data samples) and enable the distinction between a salient point and the background (area where no point of interest lies). On the other hand and depending on the application, the description is often required to provide invariance to the point of view (rotation and scale invariance) as well as variability associated to the observation process such as the noise and outliers present in the data. Therefore the desired characteristics for a descriptor depend on the target application and the characteristics of the data.

Both in 2D and 3D domains, a considerable amount of research effort has been carried on feature description. The earliest attempts at feature detection focused on detecting low-level salient patterns such as edges Canny [1986], corners Harris and Stephens [1988], Shi et al. [1994]. Inspired by these methods, modern feature description combines multiple description values to provide a detailed description. Importantly, the problem of invariance to the point of view has been investigated actively. Scale invariance, as obtained by difference of Gaussian Lindeberg [1998] and a pyramidal approach led to the popular SIFT features Lowe [1999]. RIFT descriptors Lazebnik et al. [2004] provide rotation invariance by describing each point on a set of concentric rings within which a histogram of gradient is computed. SURF description improved on RIFT and SIFT by combining both scale and rotation invariance in a computationally efficient manner by using integral images. Histograms are commonly employed in feature description as a way to obtain dimensionality reduction and robustness to noise (due to the bining process). When considering 3D data, so-called oriented descriptors base their description on the oriented normal of the local patch, effectively providing orientation invariance. Following this approach, NARF descriptor Steder et al. [2011] extended SIFT by operating a description on a star pattern around the normal of the keypoint. In order to capture relations between neighbour points on a surface, pairs of points have been employed such

as in surflet-pair histograms Wahl et al. [2003] and Fast Point Features Histograms Rusu et al. [2009] where pair-wise description provides information on the local curvature.

#### 6.3.2.2 2D histogram based description

Although lower in dimension, we expect to solve the resulting 4D matching problem in a scalable way. We choose to adopt a similar approach to the subspace analysis presented in section 6.3.1.1 by taking advantage of a rotation-invariant histogram description of each subspace and comparing it to the description of the model to infer the most likely location of the model in the scene. We assume the top of the structures to contain enough features and to be completely observed in order to enable visual recognition using a simulated view of their CAD model.

#### 6.3.2.3 Structure description

In order to account for potential yaw rotations of the model in the scene, we adopt a rotation-invariant description of the model by characterizing elevation values on $N$ complimentary circular regions as depicted in figure 6.13. So as to obtain a fine

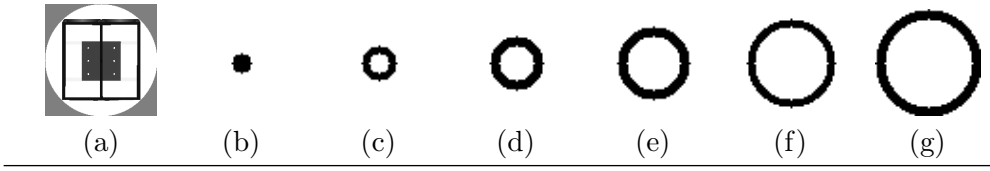



FIGURE 6.13: Model description on $N$ circular regions. The elevation map of the CAD model (a) is described using a rotation invariant descriptor computed on multiple (N = 6 in this example) complementary circular areas.

characterisation, we compute elevation histograms on each circular area, therefore reducing dimensionality through the use of histograms while preserving geometrical characteristics by separating the descriptions at each radius. The histograms are computed at a fixed depth resolution represented in $P$ bins set so as to provide specificity while allowing noise-robustness (difference between measurement and CAD model).

#### 6.3.2.4 Structure recognition

Once the object of interest described, we propose to find it in the scene by comparing it to each location in the scene map. Following the same idea used when partitioning the

3D space (see section 6.3.1.1), the scene is partitioned in multiple 2D circular patches, sampled at a sampling period along two axis orthogonal to the depth axis $T_{NE}$. Similarly to the partionning described in section 6.3.1.1, the radius of the patches $r_{patch}$ is defined by:

$$r_{patch} = r_{model} + \frac{\sqrt{2}}{2} . T_{NE} \tag{6.8}$$

with $r_{model}$ being the radius of the model as defined in section 6.3.1.1.

Using this sampling scheme and due to the completeness assumption (the top of the whole structure is visible in the data), at least one patch is guaranteed to contain the whole reconstructed structure. The recognition step then aims at finding this patch by selecting the patch that resembles the model the most. Following the description method presented in section 6.3.2.3, each patch is characterized in a rotation invariant way using an ensemble of histograms computed on complementary circular patches. Although multiple shapes can lead to the same description, the use of multiple circular patches ensures the specificity of the description while computing the histogram over the whole circular patch ensures rotation and small-noise invariance through the bining process. Since the CAD model does not initially have any absolute depth, the same depth alignment method as used in our SLAM method (see section 6.2.2.1, equation 6.1) is applied. Following the alignment of the depths of the model and the patch, the Earth Mover's Distance metric (Rubner et al. [2000]) is used to quantify their similarity on each circular region. As illustrated in figure 6.14, N distances are then obtained and concatenated in a vector. The euclidean norm of this vector is then chosen as the similarity score. The candidate patch with the lowest score is finally selected as being the best representation of the model in the scene.

#### 6.3.2.5 Model matching

Once the best patch obtained, the CAD model is finely registered in the region of recognition with a two-step registration method.

Due to the small size of the region, an exhaustive 4D ($North, East, Depth, Yaw$) is first applied with structure $Yaw$ ($\theta$ angle) varying in $[0°, 180°]$ or $[0°, 360°]$ angular intervals depending on the symmetries of the object. The depth of the model is set following the same method as described before (signed median matching described in section 6.2.2.1,

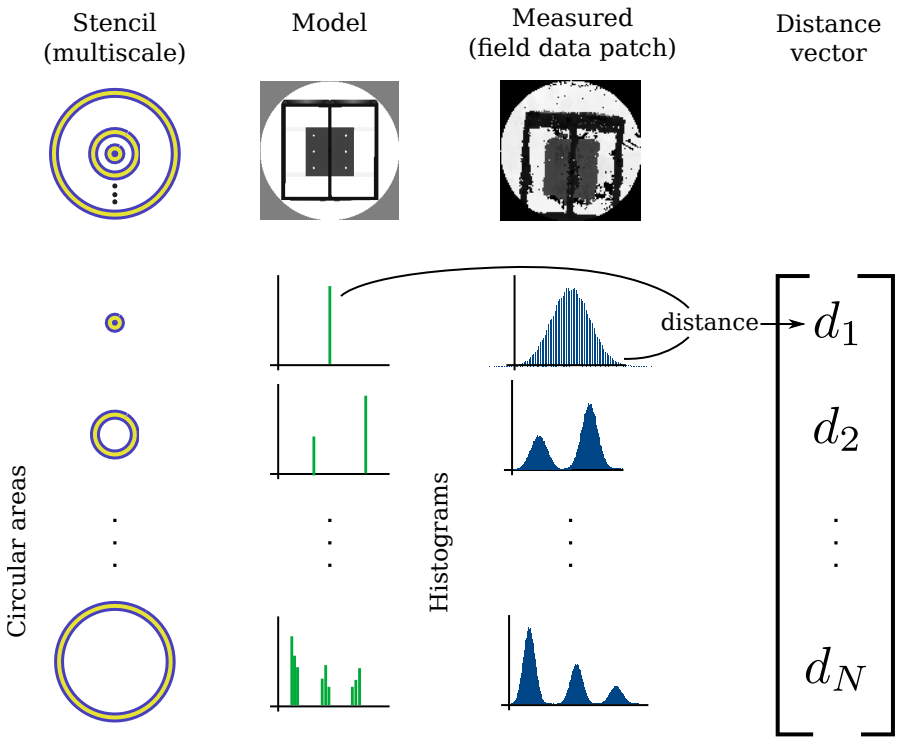



FIGURE 6.14: Histogram-based model recognition method. Each field subdivision is compared to the model using a histogram comparison for which we use Earth Mover's Distance (EMD). The distances are computed on a set of multiscale circular areas to ensure rotation invariance. The norm of the distance vector of EMD distances $[d_1,...,d_N]$ is used as a similarity score.

equation 6.1, defining $\zeta_{(\delta_N,\delta_E)}$). The optimal registration $(\delta_N^\star, \delta_E^\star, \theta^\star)$ is then selected as the minimum euclidean distance between the depths maps of the transformed model $Z_{model}$ and the field patch $Z_{map}$:

$$(\delta_N^\star,\ \delta_E^\star,\ \theta^\star) \ = \ \underset{\delta_N,\delta_E,\theta}{\arg\min} \ \| Z_{model}(i,j,\theta) - Z_{map}(i+\delta_N, j+\delta_E) + \zeta_{(\delta_N,\delta_E)} \|_2 \qquad (6.9)$$

In order to account for possible offsets in the remaining two angular dimensions (*Pitch* and *Roll*), a robust ICP-based method (Masuda et al. [1996]) discarding the furthest points (outliers) is applied.

#### 6.3.2.6 Implementation and optimizations

We implemented the 2D space model matching in C++ with the OpenCV library (Bradski et al. [2000]). The input to the algorithm is the CAD model of the structures / objects of interest, the 3D point cloud of the scene as well as the resolution of the elevation map

$\Delta_{NE}$, the width of circular regions, the depth resolution used for the histograms and the sampling period $T_{NE}$.

When a set of CAD models is available to describe the scene, the computational cost can be reduced by optimizing the modelling sequence: the structures are in this case detected in the scene sequentially and in decreasing size order, allowing the removal of large areas of the field after their detection.

Since our recognition approach is based on the minimization of a similarity score, estimating a lower bound to this score before completing the histogram comparison enables to discard wrong solution at an earlier stage. Our similarity score being an euclidean distance of a N-dimensional vector, the temporary distance given by computing the distance on $k$ elements ($1 \leq k \leq N$) naturally provides a lower bound to the similarity distance. We therefore estimate the similarity of the patches at each step by computing partial distances with increasing $k$ numbers, aborting the comparison when the distance appears larger than the current minimum distance found.

#### 6.3.2.7 Experimental results on field data

We present here results of structure recognition on the 3D reconstruction of the field of AIV trials in Fort William (see figure 6.8). The entire point cloud was converted to an elevation and the method-recognition method is applied to the entire map. The resolution of the elevation map was set to 8cm, the width of the circular patches set to 24cm, the depth resolution for the histogram bins was 1cm.

As illustrated in figure 6.15, the three CAD models (a,b,c) of the structures present on the field are first converted to elevation maps (d,e,f). The recognition algorithm then successfully finds the best patches (g,h,i) in the field. Following the detection of each model, the two-step registration technique is applied. As shown in figure 6.16-a-c, the 4D matching based on the elevation maps provides a rough registration on a fixed grid (8cm period in *North* and *East* and *Yaw* angles steps of 5°) exhibiting small angular and translation offsets. The ICP-based second registration step then corrects for the remaining gaps as can be observed in figure 6.16-d-f.

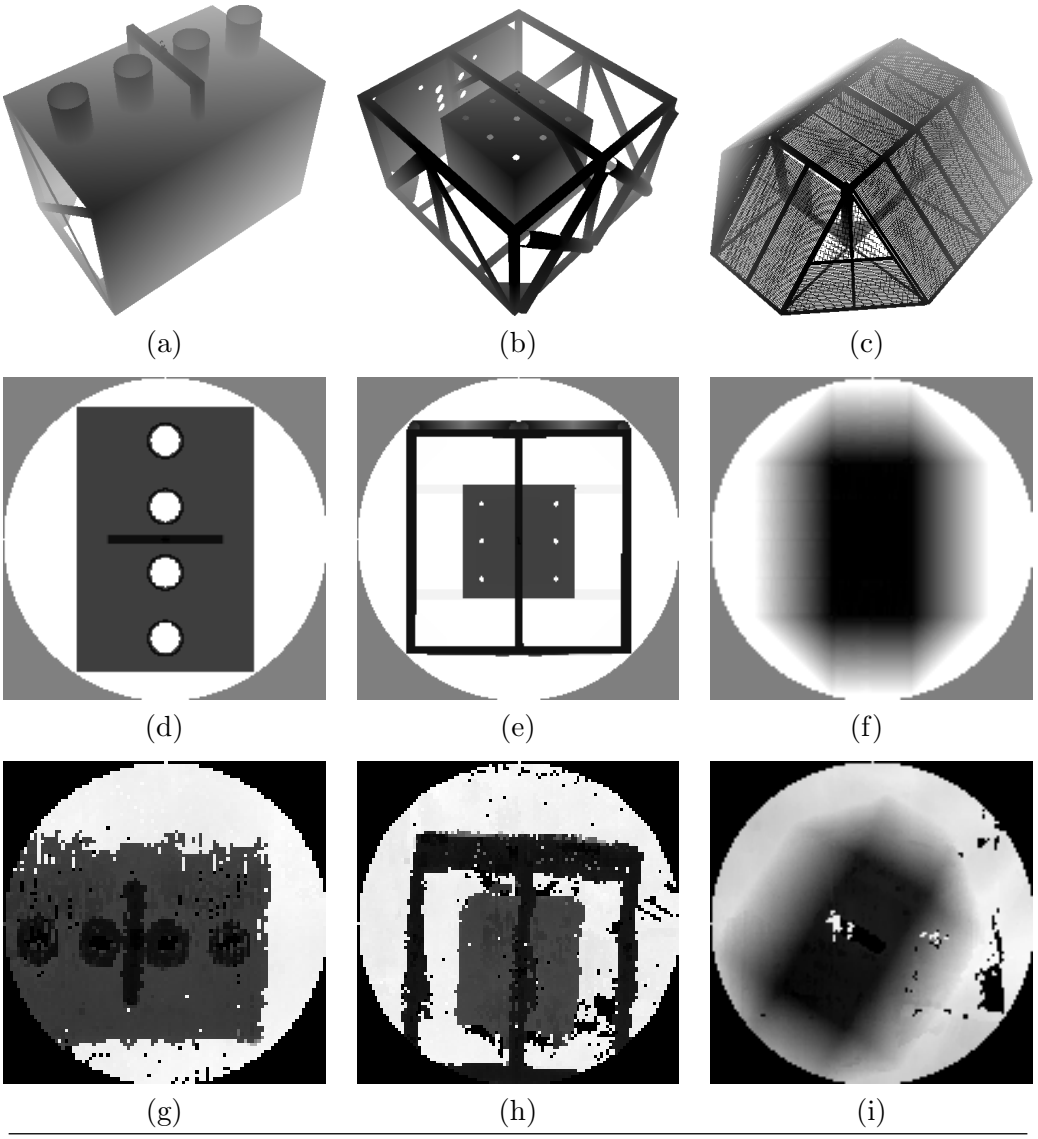



FIGURE 6.15: Model-based recognition in 2D space field results. a-b-c) The CAD models used as prior information for the object recognition step are converted to elevation maps (d-e-f) to perform a structure recognition step on the full field, leading to the selection of the three most similar patches representing the structures of interest (g-h-i).

Table 6.1 gathers distance metrics between the registered point cloud and their CAD models, exhibiting an average 40% improvement in registration when using ICP with a final average median error under 3cm.

| | **Structure 1** | **Structure 2** | **Structure 3** |
|---|---|---|---|
| **Step 1** | 8.3 (8.6) | 5.2 (7.5) | 2.7 (4.1) |
| **Step 2** | 2.4 (3.7) | 4.1 (5.4) | 1.8 (2.6) |

TABLE 6.1: Unsigned median distances (and mean distances) in cm between the reconstructed structures and their CAD models after each registration step.

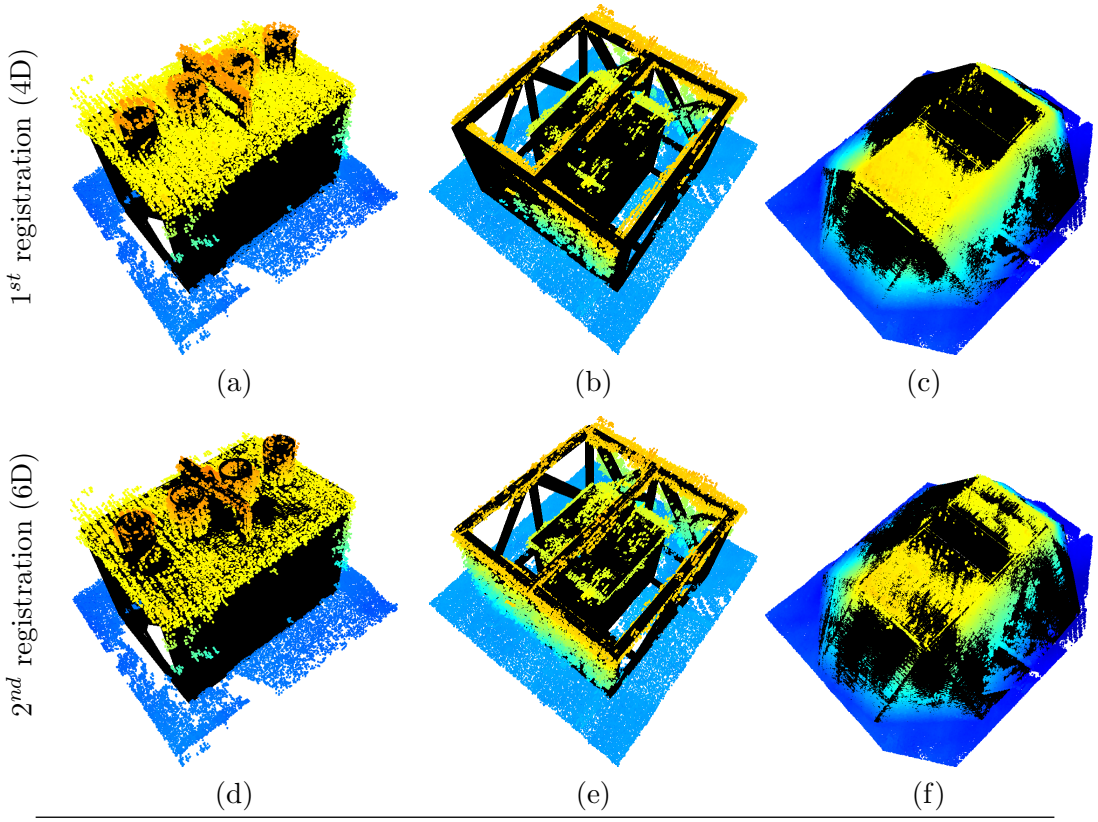



FIGURE 6.16: Two-step model registration. a-b-c) The 4D registration based on the elevation map representation provides a first rough registration of the model (represented in black) on a fixed 4D grid. d-e-f) The second registration step, based on ICP, provides a full 6D registration.

#### 6.3.2.8 Performance and limitations

The experimental results were obtained on a recent hardware configuration (Intel i7-4700MQ processor with 16GB RAM). On this platform, the identification of the two first structures was performed in 30sc at $\Delta_{NE} = 8$cm while the last and larger structure (grillage) required up to 9mn. Similarly the model registration of each structure took from 1 to 15mn. These durations depend on the size of the structure, the sampling resolution and on the symmetry to define the angular range of the first matching step. As illustrated in figure 6.17, multi-resolution tests showed the robustness of the detection and matching performances at sampling resolutions ranging from 4 to 12cm while the average computing time directly depends on the choice of the sampling period. During our experiments, the choice of the number of histogram bins did not appear determinant. Therefore and after further engineering, the choice of these parameters could be adapted to the specific situation based on the available computing resources, the size of the field, the size of the structures and the time constraints.

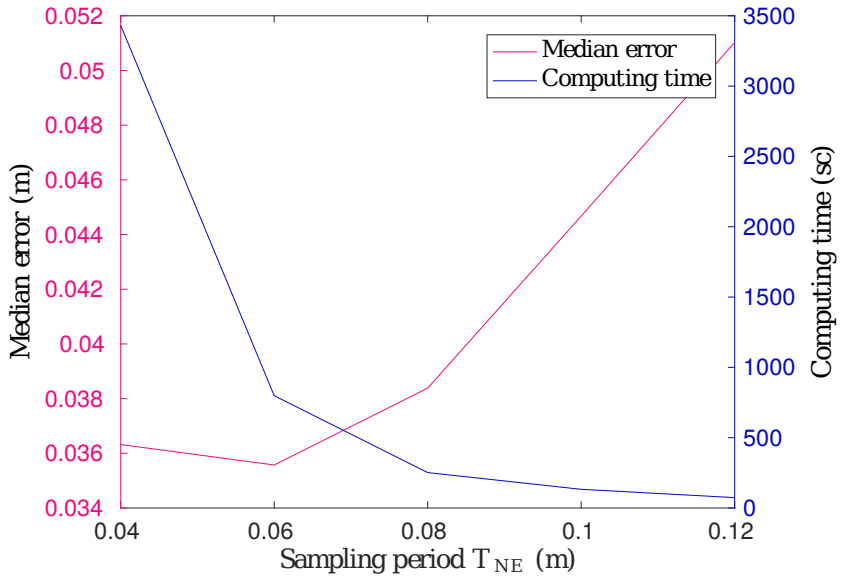



FIGURE 6.17: Influence of the sampling period $T_{NE}$ on matching error (magenta) and computing time (blue). The median distance between the model and the reconstruction after the structure recognition and the first registration step appears to be limited by the frame rate used during the survey (4cm sampling along track resolution). The computing time naturally increases with higher sampling resolutions.

Although two pipelines were present on the field of trials, the method described here only enabled to identify and match accurately the biggest one of the two (25cm radius vs 10cm radius). This is due to the high eccentricity of the pipelines resulting in circular patches mainly filled by seabed points. Therefore the percentage of points representing the structure remains very low, providing a noisy description of the object. In the case of a pipeline, a simple line detection algorithm applied to the elevation map of the field would be more adapted to the description of this specific geometry.

### 6.3.3 Applications

In comparison to the field reconstruction presented in figure 6.8, applying a model-based object recognition to the scene provides higher level, functional information on the elements of interest. Obtaining a semantic interpretation is of great interest for assisting the operation of AUV in unknown or partially unknown environments. In particular, the knowledge of semantic information paves the way to higher level applications such as landmark-based navigation, online world modelling, automated hazard detection or autonomous manipulation.

**World modelling**

Based on the recognition and registration of CAD models in the scene (both in 2D and 3D space), simple metrics can be employed to assess the fidelity of the model to the reconstructed object. Knowing the accuracy of the observation system (sensor accuracy, positioning and navigation systems), these metrics can then be used as an input to decision on the most adapted model for the scene. As an example, the matching errors presented in table 6.1 correspond to typical reconstruction error from SONAR data and are coherent with centimetre level range accuracies of the sensors employed, making the representation of these reconstructions by the models legitimate.

As a result, the 3D point cloud representation of the environment can be represented by a simplified representation based on the CAD models of the objects of interest, the remaining points representing natural features such as the seabed can then be replaced by a smooth representation such as a Poisson surface (Kazhdan et al. [2006]). In order to illustrate this, figure 6.18 presents two world models generated (b,c) generated at different levels of details from the raw 3D point cloud (a). In comparison to the initial point cloud (600MB), the first world model (figure 6.18-b) provides an accurate continuous representation of similar size (500MB) while the second world model (figure 6.18-c) exhibits a simplified representation, only requiring 5MB of storage space. In addition to visual 3D representations, a simple graph representation can be generated by simply representing the position of the objects of interest with potential connectivity between these objects, providing simple summary on the inspection results.

Through the generation of a simplified, yet accurate representation of the world, a high-level map of the world is obtained and can be integrated in both offline and online processing. A concrete application of world modelling to a real field situation comes from the difficulty to operate an AUV on old partially structured environments such as an oil field. In this situation, not only the appearance of the structures change (due to corrosion or marine growth as illustrated in figure 6.9) but the position of the objects can also differ from previous observations due to water currents affecting mobile structures (such as riser structures) and limited navigation and sensing accuracy during the previous observations. More generally and due to the frequent use of human-controlled ROVs, little information is in general available for the deployment of AUVs. As a consequence, the operation of AUVs on such fields often relies on rough world representation featuring approximate structure positions and orientations. Since the value of underwater structures and

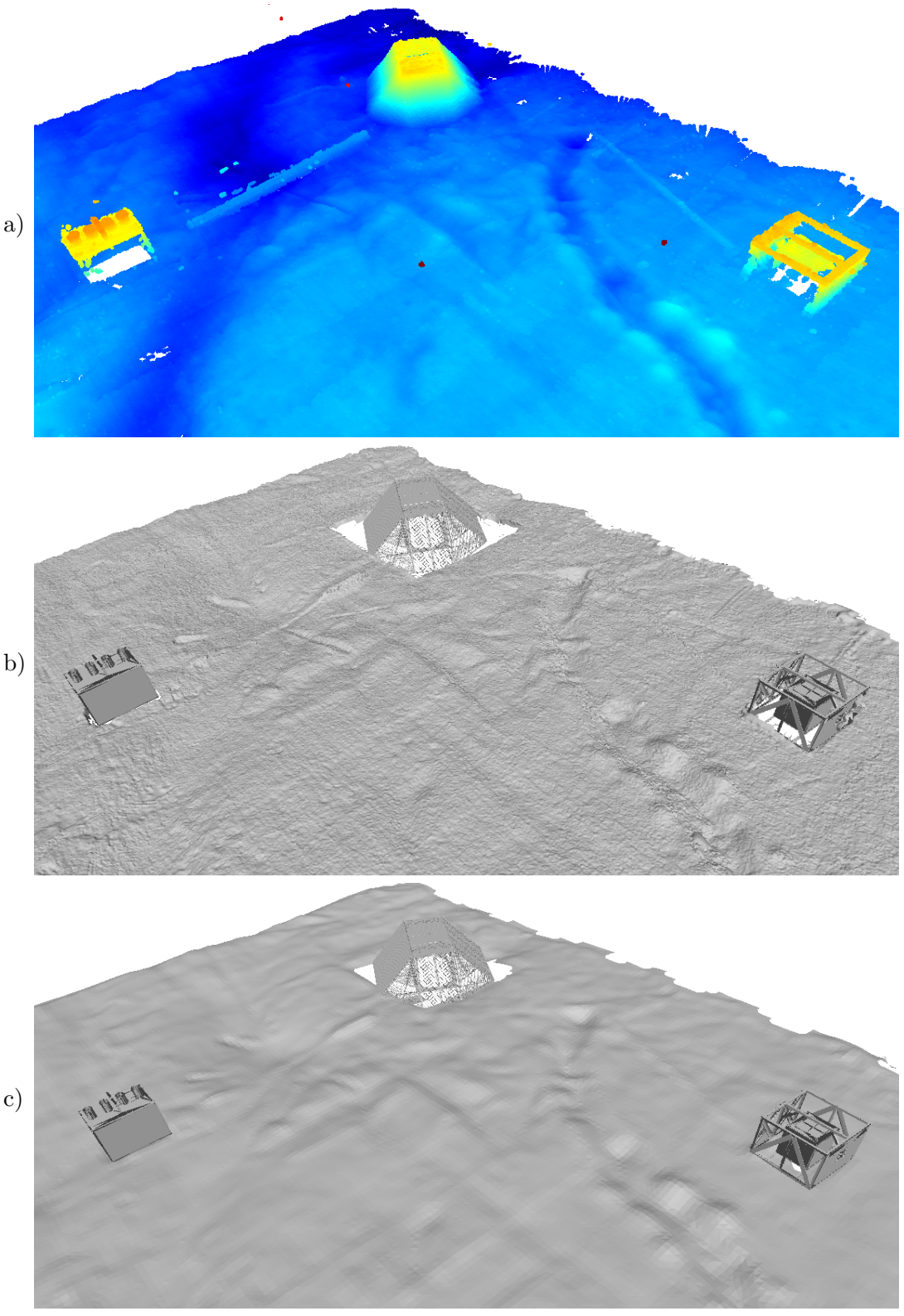



FIGURE 6.18: Illustration of World model generation. The color-coded reconstructed point cloud (6.97M points) of the full field (a) can be represented by the set of registered CAD models and a surface representation of the seabed generated at various levels of details: b) 5.95M points and 11.9M faces, c) 33k points and 65k faces.

equipment is high, safety margins are commonly required in every operation in such environment. In this situation, uncertainty on the prior model of the world makes autonomous operation significantly more complicated. In this context, the ability to generate an accurate world model that can be embedded on an resource-limited vehicle is of great interest. The operation of an AUV on a real field can therefore be made significantly safer by previously operating a complete 3D mapping of the field at a safe distance (5m in our field experiment), followed by semantic annotation such as demonstrated in our experiments.

**Navigation and mission planning**

Raw 3D point cloud representations of the environment primarily provide information on space occupancy enabling navigation correction based on the features of the terrain, path planning and collision avoidance. In comparison to this, semantic information such as knowledge on the location of a known structure on a field enables landmark-based relocalisation. Given a map of a field with a few landmarks (such as man-made structures) and a description of these landmarks (CAD model), the inevitable drift generated by dead-reckoning navigation can be compensated by the observation of these landmarks with any sensing modality. In particular, we showed that the use of standard SONAR and navigation sensors enabled the generation of a representation with enough information for landmark detection and world model generation. A top inspection with a downward-looking imaging SONAR provides sufficient information for structure matching, once the model matched, the position and orientation of the model with respect to the vehicle can be compared to the prior map enabling relocation of the vehicle with an accuracy level only depending on the matching and reconstruction method (centimetre-level).

Similarly, knowledge on the position and orientation of each object of interest can be used directly for optimized path planning. When inspecting an object, the path planning can be optimized based on the shape of the object (Xi and Shu [1999]), the sensor characteristics, the vehicle mobility and the requirements on the output data (Yang and Luo [2004]).

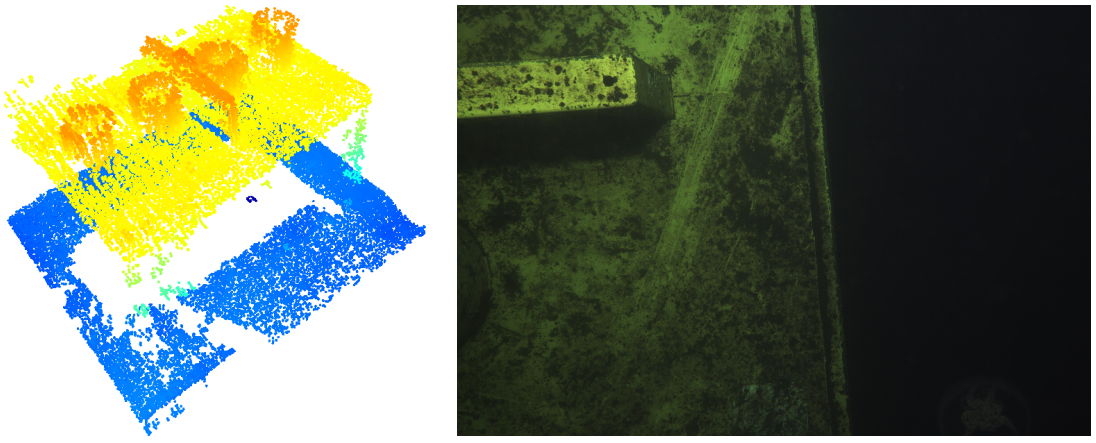
**High-level reasoning**

Based on prior information, knowledge on the nature of an object can be used for object-specific applications. In the case of our field experiments, the presence of two oil-field structures make likely the presence of connectivity between them (pipeline or

cables). Specific tasks such as leak detection can then be carried around this structure. Similarly manipulation on structure parts such as a valve can be planned based on the identification of a structure and its orientation: the location and type of a specific part can be deduced and used for model based manipulation (Hasegawa et al. [1992]).

## 6.4 Optical mapping from 3D information

Once a 3D representation of the world acquired from imaging SONAR data, the visualization of the environment is limited to information on 3D geometry of the scene whereas in practice, visual observation of the real world by human being provides colour information. In order to add optical information, we investigated the direct mapping of video images (see figure 6.19-b) on the 3D representation acquired during the AIV trials in Fort William (such as illustrated in figure 6.19-a).



(a) (b)

FIGURE 6.19: Association of a 3D representation (a) and video data (b) obtained during AIV trials using a downward-looking M12 CATHX video camera.

In the case where a video camera and a depth sensor such as a SONAR are both present on the same vehicle, it is possible to simultaneously acquire data for 3D reconstruction of the scene as well as colour information. In this situation and when the position of the two sensors on the vehicle are known, both informations can be directly co-registered. When both datasets are acquired at the same time, the knowledge of sensors offsets is sufficient to relate the data acquired with the two sensing modalities. When on the contrary, a reconstruction acquired during a previous survey is used for reference 3D representation, the presence of both sensors on a vehicle make possible the spatial

registration of the vehicle relatively to the previous representation by registering the previous 3D reconstruction to the result of the new survey. In this situation, a registration such as described in the SLAM method (see section 6.2.2.1) will be employed.

### 6.4.1 Video mapping on 3D reconstruction from SONAR

Once the 3D reconstruction of a structure (see figure 6.19-a) obtained from a series of range readings (SONAR images), the video data is simply projected onto the 3D reconstruction through the 3D raytracing of each pixel. The 3D point cloud is parsed in an Octree, providing a structure enabling look-up requests and raytracing using PCL implementation. The field dataset obtained while performing a top inspection of the brick structure was used to evaluate the performance of video mapping on the 3D reconstruction from BlueView MB2250 images. Video frames were captured at 2Hz corresponding to a motion of approximately 20cm between each frame. As illustrated in figure 6.21, the spatial resolution of the resulting representation is limited by the low resolution of the initial SONAR reconstruction. In this situation, the high resolution of the video data does not bring more spatial information but the colour information on the point cloud could be used for semantic interpretation. Assuming that the colour of the structure is known in advance and that it differs from the background colour, the colour information could be used for segmenting the man-made objects from the natural elements. In particular, the location of the structures could be detected in the colour space directly and used as a prior for model matching initialization, avoiding a costly search and matching on the point cloud of the whole field.

### 6.4.2 Model-based video mapping

As presented in section 6.3.2.7, the 3D reconstruction obtained from SONAR data exhibit enough detail for accurate model-based object recognition and model registration. Once registered to the scene, the CAD model provides reference 3D information at a much higher resolution than the SONAR 3D reconstruction. We therefore applied the video projection of the same sequence on the registered brick model (see registered model on figure 6.16-d) by discretizing the CAD model in a 1cm resolution octree. As visible in figure 6.21-a, the resulting representation exhibits a much higher level of details than the SONAR based coloured point cloud presented in figure 6.21. Due to small scale navigation

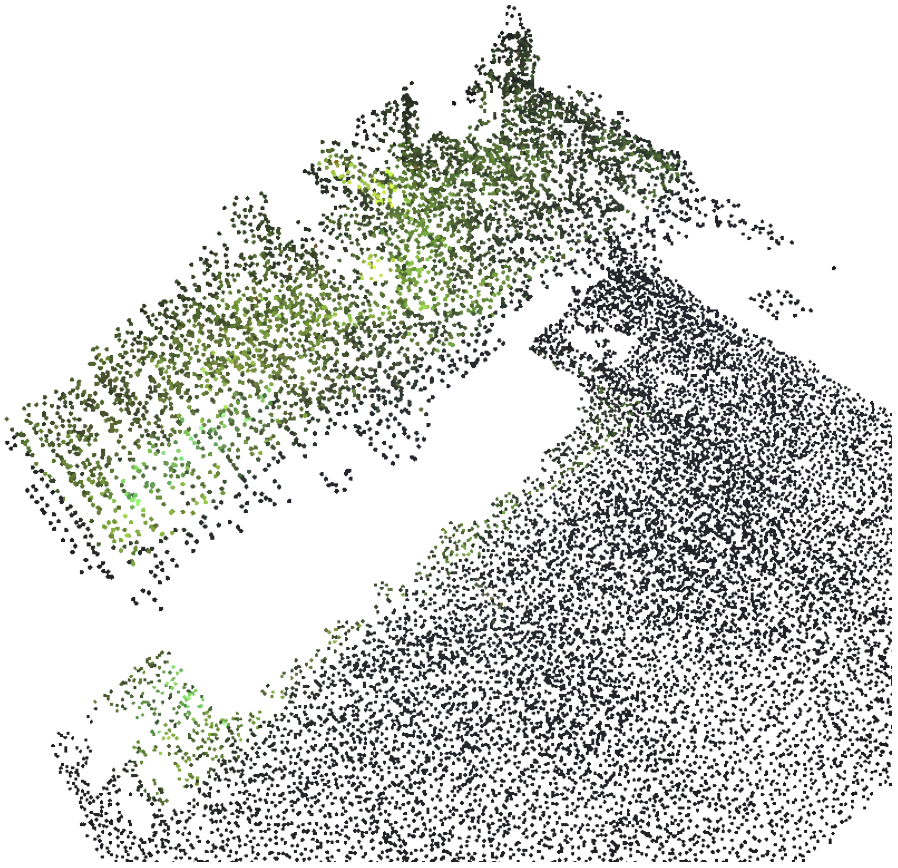
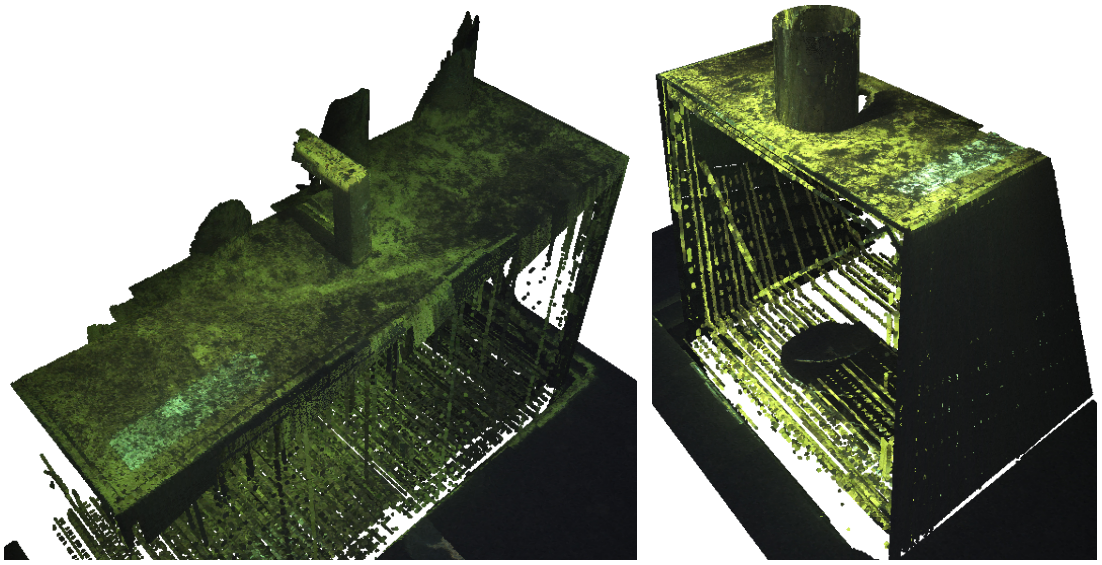


FIGURE 6.20: Illustration of a video projection on a 3D reconstruction from SONAR data. The spatial resolution is limited by the resolution of the SONAR reconstruction.

drift, a few gaps discontinuities appear (see long edge of the top of the structure). A second test sequence acquired while observing the same structure at slightly shorter range was mapped on the model (see figure 6.21-b). Due to the limited resolution of the Octree and the limitation of the raytracing algorithm in detecting collisions, a few points appear to be raytraced behind the front surface of the object, as visible at the bottom of the structure. This limitation could be overcome by raytracing directly onto the 3D surface representation (mesh) using a different raytracing implementation.

### 6.4.3 Discussion

While the spatial resolution of the result depends directly on the navigation accuracy and the initial 3D representation, the experiments made with the CAD model as reference 3D information showed the possibility of obtaining a high resolution coloured 3D representation of an underwater structure using two different sensing modalities. In this situation, the small navigation drift present in the dataset becomes visible but could



(a) (b)

FIGURE 6.21: Illustration of a video projection on a registered CAD model. The top of the structure exhibits a high level of visual features. Although unobserved, the bottom of the structure features sparse lines due to the limited resolution of the representation used for the CAD model.

be addressed by previously measuring the camera calibration matrix and co-registering the images (bundle-adjustment), effectively taking advantage visual features to reduce navigation drift when inspecting structures.

In situations where no model is available or when the geometry of the structure is expected to be significantly different from the initial model (presence of marine growth, corrosion, damaged structure, etc.), the resulting point cloud will be of low interest for direct visual inspection by a human operator but could still be used as an additional input for semantic labelling based on the colour contrast between man-made structures and typical backgrounds such as rocky or sandy seabeds.

On the contrary, when a model of the observed object is available, a low-resolution sensing modality such as a SONAR or a low resolution video camera can be used to perform object recognition and model registration. In situations where the model provides a realistic representation of the object at the time of inspection, high-resolution coloured renderings can be obtained using direct video projection. Additionally, model-based experiments showed the potential of using a high-resolution sensing modality such as a LASER, providing accurate 3D colour representations. Whether based on an accurate CAD model or using a high resolution sensor, video projection enables direct visual inspection of coloured 3D objects by a human operator, avoiding the need for frame-by-frame

inspection as commonly done. A multi-modal dataset acquired during a multi-hours field inspection could therefore be summarized into a single 3D coloured representation, considerably reducing the amount of time required for inspection by a human operator. This technique has direct applications to maintenance of man-made underwater structures such as oil field rigs.

These results points to interesting directions of research such as change detection where multiple reconstructions taken at different times could be compared both in terms of geometry and in terms of colour to detect significant changes in shape and appearance, with direct applications such as automated fault detection or mine detections.

## 6.5 Conclusions

In this chapter, we presented a few applications of 3D reconstruction from SONAR data, enabling semantic mapping of large areas with an AUV equipped of a standard 2D imaging SONAR.

We first extended the 3D reconstruction techniques to the mapping of large areas through a registration process enabling the mitigation of navigation drift. We demonstrated the benefits of this method on a medium-size field of which a large 3D point cloud representation was obtained, exhibiting better geometrical consistency in comparison to the raw reconstruction. The resulting reconstruction exhibited enough detail for being an input to navigation, path planning methods as well as semantic labelling, therefore enabling more autonomy in SONAR-equipped vehicles.

We demonstrated the possibility to obtain a semantic interpretation of the reconstructed scene through the knowledge of rough CAD model representations and two object recognition and matching techniques. Experimental results with the first method, operating matching in the 3D space, showed the recognition of simple shapes in presence of marine growth with direct applications to pipeline tracking. While the first method was limited to recognition in small scenes due to a computationally intensive implementation, the second method proposed to perform model-based structure recognition in large scenes through the use of a lightweight elevation map representation. Using this technique, experimental results on data acquired by an AUV showed the capacity to quickly identify each structure present in the scene and locate them in 6D with a centimetre-level accuracy, comparable to the SONAR sensor and local navigation accuracies.

Once the structures identified and located, we demonstrated the interest of semantic labelling for the generation of accurate 3D models of the world with semantic meaning. In comparison with traditional point clouds, semantic world models provide a lightweight and meaningful representation with applications to model-based relocalisation, optimized object inspection through model-based path planning as well as object-specific tasks such as manipulation.

We finally presented experimental results of video mapping on the 3D representations acquired of the structures of interest on the field of trials. Through this experiment, we showed the interest of performing model-based video mapping with results enabling direct

visual inspection by an operator. While restrained to situations where an accurate model of the object is available, these results suggested that the use of a high resolution sensing modality such as a laser scanner would provide similar coloured 3D representations enabling detailed visual inspection as well as more automated assessment process such as change detection between multiple observations of important parts or structures such as oil rigs or military equipments.

# Conclusion and future work

"Science may set limits to knowledge, but should not set limits to imagination."

Bertrand Russell

## 6.6 Summary

In this thesis, we pushed the state-of-the-art of 3D semantic mapping in underwater environments. Due to the limitations of optical sensing in water, we chose to focus on the extraction of 3D information using acoustic sensing.

In chapter 2, we modelled mathematically the emission, propagation and observation process of an acoustic wave in the water leading to a generic model for a single acoustic transducer. We reviewed the different types of SONAR sensors and highlighted the interest in using 2D imaging SONARs providing large footprint, short to long range sensing capability in a contained size, enabling their integration in the smallest underwater platforms. We provided a mathematical model for 2D imaging SONAR measurements, highlighting a dependency on the angle of insonification on the surface of the object resulting in intensities that depend on the shape of the object as well as the position of observation relatively to the object. In particular, we present the aperture problem inherent to the integration along the vertical aperture, resulting in the loss of information along one direction. Through the description of the imaging model and the resulting aperture problem, we show the ill-posed nature of the 3D reconstruction problem from 2D samples and justify the impossibility to recover the initial 3D representation from a single 2D image without additional information. We presented a 2D SONAR simulation

framework, integrated in UWSim simulation environment enabling the simulation of realistic scenarios such as the inspection of an oil rig. Through the use of a raytracing method and a configurable noise model, our SONAR simulation framework provided images comparable to real SONAR images such as BlueView multibeam sensors.

Once the 3D reconstruction problem introduced, we provided a review in chapter 3 on the state-of-the-art in 3D reconstruction from SONAR data. While the small footprint of pencil-beam SONARs provide low uncertainty along the vertical aperture, this comes at the cost of a low coverage or conversely the need for slower inspections (high sampling rate). When using wide-aperture multibeam sensors, the 3D reconstruction of objects has been explored through multiple methods. Shadow-based reconstructions and more generally, shape from shading methods take advantage of the assumed shading model and match it to the data either by segmentation of the image in different classes (shadow, background, object) or by an optimization. In each case, strong assumptions on the background are taken (presence of a flat seabed to observe the shadow). Non-linear methods have been explored to reconstruct the object by iteratively bounding its volume but required a large number of observations taken all around the object. Feature-based methods have been explored with little success, providing very sparse reconstructions and requiring accurate geometrical calibration. 3D reconstruction based on the prior knowledge of a CAD model has been explored but remain of limited usage in unknown or partially-unknown environments. The use of multiple sensors, including the combination of optical and acoustic sensing modalities showed limitations in the calibration process due to differences in resolution and sensor noise. 3D reconstruction from acoustic stereo imaging has been investigated but exhibited reconstructions of limited accuracy. 3D SONAR sensors have been successfully developed, showing good reconstructions results but to this date they remain bulky and expensive, prohibiting their use on small to medium-size underwater platforms.

Following the review of 3D reconstruction methods, we presented in chapter 4 our first 3D reconstruction method based on a non-linear formulation enabling the reconstruction through the observation of empty spaces. In comparison to state-of-the-art methods, this method does not operate on shadow information hence does not require the objects to be placed on a flat surface such as a seabed. In addition to this, the method applies to sensors of any vertical aperture and no strong constraint is required on the sampling scheme. This so-called space carving technique enables an online reconstruction of the scene by

maintaining a temporary map and iteratively adding new observations. An occlusion resolution post-processing is then applied to provide the final map by retaining only the observed part of the scene (front surface of the objects). Through both a qualitative and quantitative analysis, we demonstrated the performance of this method on both simulated and real data with centimetre-level 3D resolution with both pencil-beam and wide-beam sensors. Using a sensor noise modelling step in open water, resilience to noise is achieved at diverse SNR levels, providing constant reconstruction error and outlier rate (5%) but a decreasing coverage (down to 10% at a SNR of 5dB). Importantly, this reconstruction approach requires a low computational cost (500ms per update on a single-threaded configuration) but required up to 16 Gb in the current implementation and for a large datasets (1000 samples).

Inspired by SAS techniques, we presented in chapter 5 a new formulation of the 3D reconstruction problem from a set of samples acquired along a straight line at regular sampling steps. In comparison to state-the-art reconstruction methods and similarly to the carving method, the deconvolution approach does not require large rotations around the object or shadow information. Based on this sampling assumption, we reformulated the reconstruction problem as a spatially-varying blind deconvolution. In order to solve this notoriously complicated problem, we adopted a model-to-data matching approach and a non-negative least square optimization implementation. We investigated the interest of adding regularization to the optimization process. We presented reconstruction results on simulated and real data and performed a quantitative analysis exhibiting an increase in surface coverage when using the deconvolution approach over the space carving method. In spite of the sampling constraint, field experiments demonstrated the ability to use the deconvolution method in conditions where the sampling scheme is not strictly followed with equivalent or better results than the space carving method. The interest of added regularization was exhibited on real data as well as limitations on the choice of the regularization parameter and unneeded sparsity promotion. A comparative study of our two reconstruction methods was presented highlighting the practical differences in usage: the space carving method offers unconstrained sampling scheme and sample per sample reconstruction while the deconvolution formulation required stricter sampling and the provision of all the samples at once (batch processing). While leading to comparable reconstruction errors, the two methods differ on the surface coverage (up to twice as large) yielded and the outliers structure. Practical differences were also shown due to

the low computing resources requirements of the deconvolution approach (typically 2Gb) but with a lower scalibility (large number of samples require long processing time due to the complexity of large optimization problems). We finally highlighted the interest in using the simulation environment as a pre-mission tool to optimize the sensing strategy and choose the most adequate sensor relatively to a specific situation. In particular, our quantitative analysis exhibited the interest in using commonly embedded wide-aperture SONARs to provide 3D reconstructions with higher surface coverage than pencil-beam sensors (typically 50% increase between $1^\circ$ and $10^\circ$ sensors), traditionally added on the vehicle for specific applications.

The results presented in chapter 4 and 5 exhibiting enough detail for visual recognition of the objects, we presented in chapter 6 our efforts toward the extraction of semantic information from the reconstruction of a 3D underwater scene. We first extended our 3D reconstruction work to the mapping of large area by mitigating the navigation drift observed after a few tens metres motion by applying a registration of reconstructed swathes. We demonstrated this SLAM technique on a medium-size field ($50 \times 34$ metres) and exhibited the improvement in geometrical accuracy on a few visual landmarks (oil-field structures) present on the field in spite of significant navigation drift (38cm in North and 18cm in East). We presented an inital experimental approach to structure recognition in SONAR point clouds using direct CAD model matching (ICP and Go-ICP) and exhibited their prohibitive computational cost due to the high dimensionality of the problem (20mn matching time for a small scene). Based on the full field reconstruction operated through the bundle-adjustment technique, we demonstrated the first model-based underwater semantic mapping enabling the recognition and localisation of three structures of interest present on the field with a processing that requires a few minutes only. The median registration error obtained over three structures was under 3cm and included reconstruction and measurement uncertainty as well as low-scale navigation drift. Based on this structure recognition, we showed that lightweight yet geometrically accurate and semantically meaningful maps could be generated with a representation memory footprint reducing from 600MB to 5MB. This world modelling method enables to provide a first representation of unknown environments, frequently encountered when operating in offshore environments such as oil rigs. In addition to this, these semantic maps are of use for both offline mission planning and online navigation and path planning. Taking advantage of the accurate registration of the models of the structure, we finally

showed the potential of video mapping on CAD model to obtain dense coloured 3D representations at much higher resolution (cm-level) than the initial 3D representation obtained from SONAR data (3cm resolution).

## 6.7 Future work

In chapter 2, we presented our 2D SONAR simulation results. Due to their intensive computing resource usage, multipath returns were not included in our simulation. As a result the simulation of objects in closed environments such as a water tank appears unrealistic. In addition to this, the beamforming process was not modelled as such. As a result a visible difference in sharpness appears between the simulated data and typical real images. In general the modelling of physical process is always limited to a certain accuracy and needs to be traded off with resource usage. Recently, machine learning techniques (Shrivastava et al. [2016]) have shown their interest for bridging the gap between mathematical models and real data, enabling the generation of sensor or environment-specific realistic data in a simulated environment. We believe that such techniques could be applied to our simulation framework.

Due to the nature of the space carving method presented in chapter 3, the quality of the reconstruction is sensitive to the accuracy of the positioning: the observation of an empty space wrongly located due to navigation inaccuracy results in wrongly deleting 3D occupied points. In this situation, the integration of an estimate of navigation accuracy such as typically provided by Kalman filters could prevent this phenomenon. Similarly, adding regularity constraints when performing the occlusion resolution could prevent the generation of holes in surfaces observed at grazing angle. Finally, the carving algorithm being effectively a volume bounding technique could rely on a volume representation rather than a 3D point grid. Direct meshes could potentially be generated instead of using suboptimal 3D grids.

The deconvolution method presented in chapter 4 is limited to rectilinear and uniform sampling along the direction of uncertainty. A formulation allowing more general motions would benefit to its use on an AUV with trajectories constrained by the environment.

While the results of our SLAM implementation presented in chapter 5 exhibited good accuracy, a comparison to USBL localisation on a few scenario would enable to assess the

robustness of the technique. In particular, a study of the minimum overlap between each consecutive swathes and the minimum amount of features on the seabed (relatively to the sensor accuracy) would be beneficial to a further integration on an operational AUV.

Based on our first semantic mapping results on the field of trials presented in chapter 5, there is an interest in expanding the object recognition to a wider range of submerged objects. Oil and gas offshore field typically feature tenths of different structures, mine countermeasure operations aim at detecting mines of various shapes. In this situation, the ability to operate recognition from a large set of models is of direct interest for automated classification. In addition to variation in shapes, underwater structures can be found in various sizes: pipelines, cables as well as underwater mines need to be detected in a scale invariant way. There is therefore an interest to extend our work to the scale-invariant recognition of objects from a large catalogue of models. In addition to larger scale recognition, the recognition of structure parts such as valves and switches would enable autonomous manipulation and detailed inspection. Taking advantage of accurate sensing modalities such as LIDARs would therefore enable a multi-level semantic mapping. In each of these situations, the online semantic labelling enables online or offline automated decision and manipulation, reducing considerably both the workload and stress of human operators as well as operation time. Building up on our first results, we therefore plan to extend our research by the use of different sensing modalities and a wider range of 3D models.

In chapter 5 we presented our initial results on model-based video mapping. Exploring the video mapping possibilities on a finer 3D representation such as obtained from underwater LIDARs is of interest to provide photorealistic 3D representation for visual inspection and recognition of underwater structures by human operators. Conversely, the use of both geometrical features and video information could be explored for correcting the local navigation drift occurring when inspecting a structure.

# Examples of commercial 2D SONARs specifications

| | **BlueView P900** | **Aris Explorer 3000** | **Imagenex 837A Delta T** | **Tritech Gemini 720ik** |
|---|---|---|---|---|
| **Frequency (kHz)** | 900 | 3000 | 260 | 720 |
| **Vertical beamwidth (deg)** | 20 | 14 | 10 | 20 |
| **Horizontal beamwidth (deg)** | 1 | 0.25 | 0.75 | 1 |
| **FOV (deg)** | 90 | 30 | 120 | 120 |
| **Min range (m)** | 2 | 0.2 | 0.5 | 0.2 |
| **Max range (m)** | 100 | 15 | 150 | 120 |
| **Max range resolution (mm)** | 13 | 3 | 10 | 8 |
| **Depth rating (m)** | 1000 | 300 | 1000 | 350 |
| **Power (W)** | 20 | 20 | 5 | 27 |

[Specifications of typical 2D wide-aperture SONARs]

TABLE 2: Specifications of typical 2D wide-aperture SONARs

# Model-based field structure 3D recognition and matching

Using the method described in section 6.3.1, the three CAD models were used as an input to the algorithm as well as the field reconstruction of Fort William AIV trials presented in section 6.2.2.2-figure 6.8. After downsampling the input data at 3cm resolution, the ICP-based detection respectively took 29mn20sc for the box structure and 2h for the grillage structure with successful registration. In case of the brick structure, the ICP-base registration did not converge to the position of the brick structure, rather the box structure due to a local minimum.

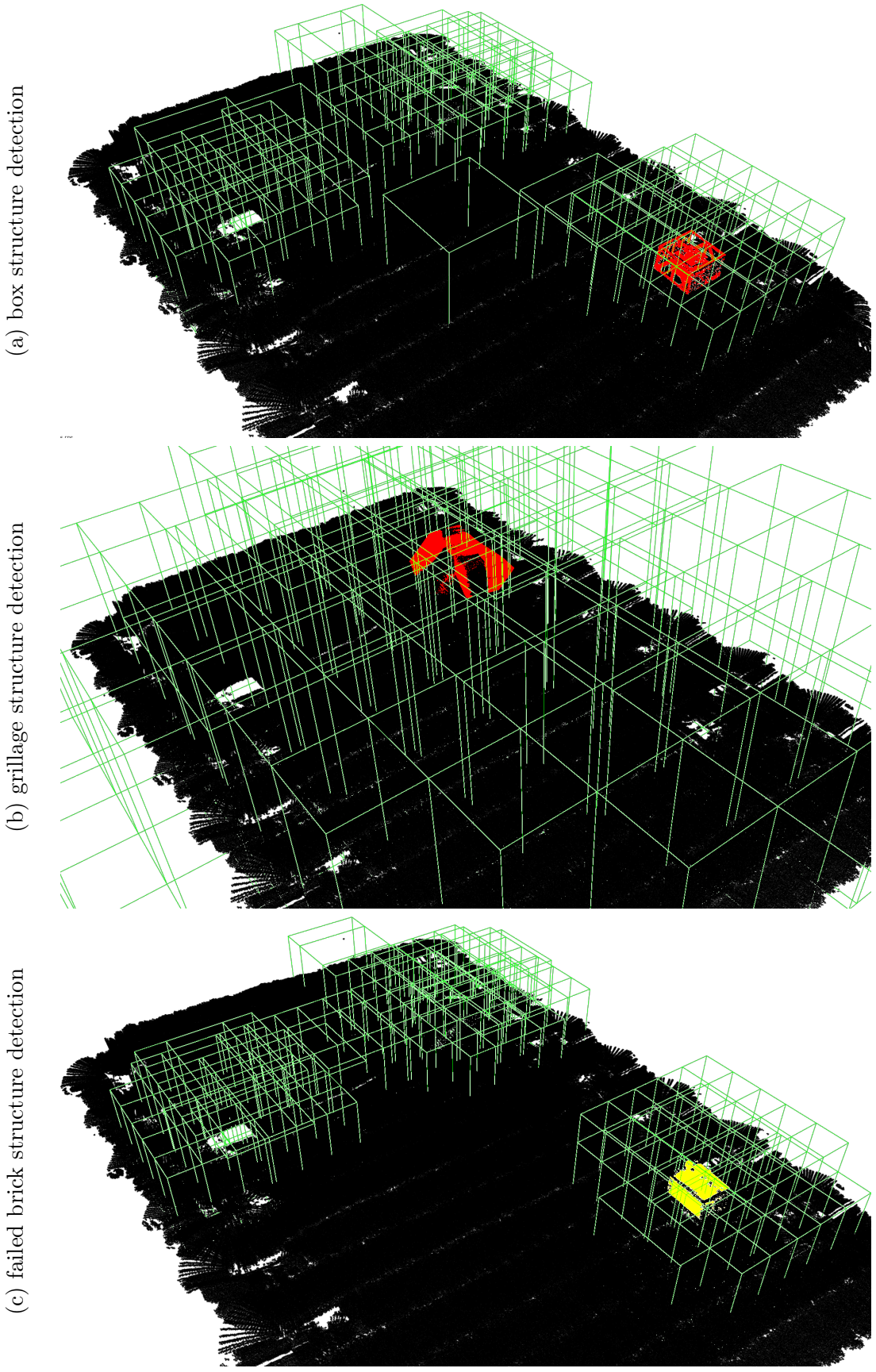



FIGURE 22: Structure detection on FW AIV trials field reconstruction using ICP-based recognition algorithm.

# Bibliography


Abbott, J. G. and Thurstone, F. (1979). Acoustic speckle: Theory and experimental analysis. *Ultrasonic imaging*, 1(4):303–324.

am Ende, B. A. (2001). 3d mapping of underwater caves. *IEEE Computer Graphics and Applications*, 21(2):14–20.

Amiri-Simkooei, A. R., Snellen, M., and Simons, D. G. (2011). Principal component analysis of single-beam echo-sounder signal features for seafloor classification. *IEEE Journal of Oceanic Engineering*, 36(2):259–272.

Assalih, H. et al. (2013). *3D reconstruction and motion estimation using forward looking sonar*. PhD thesis, Heriot-Watt University.

Auran, P. G. and Malvig, K. E. (1996). Real-time extraction of connected components in 3-d sonar range images. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 580–585. IEEE.

Aykin, M. D. and Negahdaripour, S. (2013). Forward-look 2-d sonar image formation and 3-d reconstruction. In *Oceans-San Diego, 2013*, pages 1–10. IEEE.

Aykin, M. D. and Negahdaripour, S. (2016). Three-dimensional target reconstruction from multiple 2-d forward-scan sonar views by space carving. *IEEE Journal of Oceanic Engineering*.

Babaee, M. and Negahdaripour, S. (2015). 3-d object modeling from 2-d occluding contour correspondences by opti-acoustic stereo imaging. *Computer Vision and Image Understanding*, 132:56–74.

Bagnitsky, A., Inzartsev, A., Pavin, A., Melman, S., and Morozov, M. (2011). Side scan sonar using for underwater cables & pipelines tracking by means of auv. In *2011 IEEE*

*Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*.

Barkby, S., Williams, S. B., Pizarro, O., and Jakuba, M. V. (2011). A featureless approach to efficient bathymetric slam using distributed particle mapping. *Journal of Field Robotics*, 28(1):19–39.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. *Computer vision–ECCV 2006*, pages 404–417.

Belcher, E., Hanot, W., and Burch, J. (2002). Dual-frequency identification sonar (didson). In *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on*, pages 187–192. IEEE.

Belcher, E. O., Lynn, D. C., Dinh, H. Q., and Laughlin, T. J. (1999). Beamforming and imaging with acoustic lenses in small, high-frequency sonars. In *OCEANS'99 MTS/IEEE. Riding the Crest into the 21st Century*, volume 3, pages 1495–1499. IEEE.

Bell, J. M. (1997). Application of optical ray tracing techniques to the simulation of sonar images. *Optical Engineering*, 36(6):1806–1813.

Bell, J. M., Chantler, M. J., and Wittig, T. (1999). Sidescan sonar: a directional filter of seabed texture? *IEE Proceedings-Radar, Sonar and Navigation*, 146(1):65–72.

Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics.

Bichucher, V., Walls, J. M., Ozog, P., Skinner, K. A., and Eustice, R. M. (2015). Bathymetric factor graph slam with sparse point cloud alignment. In *OCEANS'15 MTS/IEEE Washington*, pages 1–7. IEEE.

Billingsley, J. and Kinns, R. (1976). The acoustic telescope. *Journal of Sound and Vibration*, 48(4):485–510.

Bingham, B., Foley, B., Singh, H., Camilli, R., Delaporta, K., Eustice, R., Mallios, A., Mindell, D., Roman, C., and Sakellariou, D. (2010). Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. *Journal of Field Robotics*, 27(6):702–717.

Birk, A., Pathak, K., Vaskevicius, N., Pfingsthorn, M., Poppinga, J., and Schwertfeger, S. (2010). Surface representations for 3d mapping. *KI-Künstliche Intelligenz*, 24(3):249–254.

Bradski, G. et al. (2000). The opencv library. *Doctor Dobbs Journal*, 25(11):120–126.

Brahim, N., Guériot, D., Daniel, S., and Solaiman, B. (2011). 3d reconstruction of underwater scenes using didson acoustic sonar image sequences through evolutionary algorithms. In *OCEANS, 2011 IEEE-Spain*, pages 1–6. IEEE.

Brahim, N., Guériot, D., Daniely, S., and Solaiman, B. (2010). 3d reconstruction of underwater scenes using image sequences from acoustic camera. In *OCEANS 2010 IEEE-Sydney*, pages 1–8. IEEE.

Brown, C. J. and Blondel, P. (2009). Developments in the application of multibeam sonar backscatter for seafloor habitat mapping. *Applied Acoustics*, 70(10):1242–1247.

Brown, C. J., Smith, S. J., Lawton, P., and Anderson, J. T. (2011). Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, 92(3):502–520.

Bülow, H. and Birk, A. (2011). Spectral registration of noisy sonar data for underwater 3d mapping. *Autonomous Robots*, 30(3):307–331.

Callow, H. J. (2003). Signal processing for synthetic aperture sonar image enhancement.

Cannon, M. (1976). Blind deconvolution of spatially invariant image blurs with phase. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(1):58–63.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698.

Chan, T. F. and Wong, C.-K. (1998). Total variation blind deconvolution. *IEEE transactions on Image Processing*, 7(3):370–375.

Chen, V. L., Batalin, M. A., Kaiser, W. J., and Sukhatme, G. (2008). Towards spatial and semantic mapping in aquatic environments. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 629–636. IEEE.

Coiras, E., Petillot, Y., and Lane, D. M. (2007). Multiresolution 3-d reconstruction from side-scan sonar images. *IEEE Transactions on Image Processing*, 16(2):382–390.

Coiras, E., Ramirez-Montesinos, A., and Groen, J. (2009). Gpu-based simulation of side-looking sonar images. In *OCEANS 2009-EUROPE*, pages 1–6. IEEE.

Collier, J. and Humber, S. (2007). Time-lapse side-scan sonar imaging of bleached coral reefs: a case study from the seychelles. *Remote Sensing of Environment*, 108(4):339–356.

Cook, C. (2012). *Radar signals: An introduction to theory and application*. Elsevier.

Craighead, J., Murphy, R., Burke, J., and Goldiez, B. (2007). A survey of commercial & open source unmanned vehicle simulators. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 852–857. IEEE.

Davis, A. and Lugsdin, A. (2005). High speed underwater inspection for port and harbour security using coda echoscope 3d sonar. In *OCEANS, 2005. Proceedings of MTS/IEEE*, pages 2006–2011. IEEE.

Dippé, M. A. and Wold, E. H. (1985). Antialiasing through stochastic sampling. *ACM Siggraph Computer Graphics*, 19(3):69–78.

Drost, B., Ulrich, M., Navab, N., and Ilic, S. (2010). Model globally, match locally: Efficient and robust 3d object recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 998–1005. Ieee.

Durá, E., Bell, J., and Lane, D. (2004). Reconstruction of textured seafloors from side-scan sonar images. *IEE Proceedings-Radar, Sonar and Navigation*, 151(2):114–126.

Eustice, R., Camilli, R., and Singh, H. (2005). Towards bathymetry-optimized doppler re-navigation for auvs. In *OCEANS, 2005. Proceedings of MTS/IEEE*, pages 1430–1436. IEEE.

Fairfield, N., Kantor, G., and Wettergreen, D. (2007). Real-time slam with octree evidence grids for exploration in underwater tunnels. *Journal of Field Robotics*, 24(1-2):03–21.

Farr, H. K. (1980). Multibeam bathymetric sonar: sea beam and hydro chart. *Marine Geodesy*, 4(2):77–93.

Ferreira, F., Machado, D., Ferri, G., Dugelay, S., and Potter, J. (2016). Underwater optical and acoustic imaging: A time for fusion? a brief overview of the state-of-the-art. In *OCEANS 2016 MTS/IEEE Monterey*, pages 1–6. IEEE.

Folkesson, J., Leonard, J., Leederkerken, J., and Williams, R. (2007). Feature tracking for underwater navigation using sonar. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 3678–3684. IEEE.

Gauthier, S., Boisclair, D., and Legendre, P. (1997). Evaluation of a variable angle scanning method to estimate relative abundance and distribution of fish using a single-beam echosounder in shallow lakes. *Journal of fish biology*, 50(1):208–221.

Gerlotto, F., Georgakarakos, S., and Eriksen, P. K. (2000). The application of multibeam sonar technology for quantitative estimates of fish density in shallow water acoustic surveys. *Aquatic Living Resources*, 13(5):385–393.

Gerlotto, F., Soria, M., and Fréon, P. (1999). From two dimensions to three: the use of multibeam sonar for a new approach in fisheries acoustics. *Canadian Journal of Fisheries and Aquatic Sciences*, 56(1):6–12.

Goddard, R. P. (2008). The sonar simulation toolset, release 4.6: Science, mathematics, and algorithms. Technical report, DTIC Document.

Gott, R. M., Martinez, A. B., and Bourgeois, B. S. (1993). Application of backscatter models to sass. In *Southeastcon'93, Proceedings., IEEE*, pages 8–p. IEEE.

Graham, P. and Nelson, B. (1998). Fpga-based sonar processing. In *Proceedings of the 1998 ACM/SIGDA sixth international symposium on Field programmable gate arrays*, pages 201–208. ACM.

Grasmueck, M., Eberli, G. P., Viggiano, D. A., Correa, T., Rathwell, G., and Luo, J. (2006). Autonomous underwater vehicle (auv) mapping reveals coral mound distribution, morphology, and oceanography in deep water of the straits of florida. *Geophysical Research Letters*, 33(23).

Griffiths, H., Rafik, T., Meng, Z., Cowan, C., Shafeeu, H., and Anthony, D. (1997). Interferometric synthetic aperture sonar for high resolution 3-d mapping of the seabed. *IEE Proceedings-Radar, Sonar and Navigation*, 144(2):96–103.

Groen, J. (2006). Adaptive motion compensation in sonar array processing.

Gueriot, D. (2000). Bathymetric and side-scan data fusion for sea-bottom 3d mosaicing. In *OCEANS 2000 MTS/IEEE Conference and Exhibition*, volume 3, pages 1663–1668. IEEE.

Gueriot, D., Sintes, C., and Garello, R. (2007). Sonar data simulation based on tube tracing. In *OCEANS 2007-Europe*, pages 1–6. IEEE.

Guerneve, T. and Petillot, Y. (2015). Underwater 3d reconstruction using blueview imaging sonar. In *OCEANS 2015-Genova*, pages 1–7. IEEE.

Guerneve, T., Subr, K., and Petillot, Y. (2017a). Cad-model-based 3d video mapping. In *OCEANS–Anchorage, 2017*, pages 1–5. IEEE.

Guerneve, T., Subr, K., and Petillot, Y. (2017b). Underwater 3d structures as semantic landmarks in sonar mapping. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pages 614–619. IEEE.

Guerneve, T., Subr, K., and Petillot, Y. (2018). Three-dimensional reconstruction of underwater objects using wide-aperture imaging sonar. *Journal of Field Robotics*.

Hagen, P. E., Storkersen, N., Vestgard, K., and Kartvedt, P. (2003). The hugin 1000 autonomous underwater vehicle for military applications. In *OCEANS 2003. Proceedings*, volume 2, pages 1141–1145. IEEE.

Hansen, R. and Andersen, P. (1996). A 3d underwater acoustic camera—properties and applications. In *Acoustical Imaging*, pages 607–611. Springer.

Hansen, R. K., Castellani, U., Murino, V., Fusiello, A., Puppo, E., Papaleo, L., Pittore, M., Gobbi, M., Bisone, L., Kleppe, K., et al. (2005). Mosaicing of 3d sonar data sets-techniques and applications. In *OCEANS, 2005. Proceedings of MTS/IEEE*, pages 2326–2333. IEEE.

Harmeling, S., Michael, H., and Schölkopf, B. (2010). Space-variant single-image blind deconvolution for removing camera shake. In *Advances in Neural Information Processing Systems*, pages 829–837.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer.

Hartley, J., Watson, T., et al. (1993). Investigation of a north sea oil platform drill cuttings pile. In *Offshore Technology Conference*. Offshore Technology Conference.

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

Hasegawa, T., Suehiro, T., and Takase, K. (1992). A model-based manipulation system with skill-based execution. *IEEE Transactions on Robotics and Automation*, 8(5):535–544.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Horner, D., McChesney, N., Masek, T., and Kragelund, S. (2009). 3d reconstruction with an auv mounted forward-looking sonar. Technical report, DTIC Document.

Hover, F. S., Eustice, R. M., Kim, A., Englot, B., Johannsson, H., Kaess, M., and Leonard, J. J. (2012). Advanced perception, navigation and planning for autonomous in-water ship hull inspection. *The International Journal of Robotics Research*, 31(12):1445–1464.

Huang, S.-q., Liu, D.-z., Gao, G.-q., and Guo, X.-j. (2009). A novel method for speckle noise reduction and ship target detection in sar images. *Pattern Recognition*, 42(7):1533–1542.

Huang, T. A. and Kaess, M. (2015). Towards acoustic structure from motion for imaging sonar. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 758–765. IEEE.

Hurtós, N., Cufi, X., and Salvi, J. (2010). Calibration of optical camera coupled to acoustic multibeam for underwater 3d scene reconstruction. In *OCEANS 2010 IEEE-Sydney*, pages 1–7. IEEE.

Hurtós, N., Nagappa, S., Palomeras, N., and Salvi, J. (2014). Real-time mosaicing with two-dimensional forward-looking sonar. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 601–606. IEEE.

Hutin, E., Simard, Y., and Archambault, P. (2005). Acoustic detection of a scallop bed from a single-beam echosounder in the st. lawrence. *ICES Journal of Marine Science: Journal du Conseil*, 62(5):966–983.

Isar, A., Moga, S., and Isar, D. (2005). A new method for denoising sonar images. In *Signals, Circuits and Systems, 2005. ISSCS 2005. International Symposium on*, volume 2, pages 469–472. IEEE.

Jaffe, J. S., Reuss, E., McGehee, D., and Chandran, G. (1995). Ftv: a sonar for tracking macrozooplankton in three dimensions. *Deep Sea Research Part I: Oceanographic Research Papers*, 42(8):1495–1512.

Johnson, A. E. and Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):433–449.

Joshi, N., Kang, S. B., Zitnick, C. L., and Szeliski, R. (2010). Image deblurring using inertial measurement sensors. In *ACM Transactions on Graphics (TOG)*, volume 29, page 30. ACM.

Josserand, T. and Wolley, J. (2011). A miniature high resolution 3-d imaging sonar. *Ultrasonics*, 51(3):275–280.

Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). Poisson Surface Reconstruction. In Sheffer, A. and Polthier, K., editors, *Symposium on Geometry Processing*. The Eurographics Association.

Kenny, A., Cato, I., Desprez, M., Fader, G., Schüttenhelm, R., and Side, J. (2003). An overview of seabed-mapping technologies in the context of marine habitat classification. *ICES Journal of Marine Science: Journal du Conseil*, 60(2):411–418.

Kim, S., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D. (2007). A method for large-scale l1-regularised least squares problems with applications in signal processing and statistics. *IEEE Journal on Selected Topics in Signal Processing*, 1:606–617.

Kino, G. S. (1987). *Acoustic waves: devices, imaging and analog signal processing*. Number 43 KIN.

Kinsey, J. C., Eustice, R. M., and Whitcomb, L. L. (2006). A survey of underwater vehicle navigation: Recent advances and new challenges. In *IFAC Conference of Manoeuvering and Control of Marine Craft*, volume 88.

Krishnan, D., Tay, T., and Fergus, R. (2011). Blind deconvolution using a normalized sparsity measure. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 233–240. IEEE.

Kwangmoo Koh, S.-J. K. and Boyd, S. (2008). Simple matlab solver for l1-regularized least squares problems. https://stanford.edu/~boyd/l1_ls/. [Online; accessed 25-April-2017].

Lab, I. (2012). Uwsim - the underwater simulator.

Langer, D. and Hebert, M. (1991). Building qualitative elevation maps from side scan sonar data for autonomous underwater navigation. In *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on*, pages 2478–2483. IEEE.

Lanzoni, J. and Weber, T. (2010). High-resolution calibration of a multibeam echo sounder. In *OCEANS 2010*, pages 1–7. IEEE.

Latombe, J.-C. (2012). *Robot motion planning*, volume 124. Springer Science & Business Media.

Lawson, C. L. and Hanson, R. J. (1974). *Solving least squares problems*, volume 161. SIAM.

Lazebnik, S., Schmid, C., and Ponce, J. (2004). Semi-local affine parts for object recognition. In *British Machine Vision Conference (BMVC'04)*, pages 779–788. The British Machine Vision Association (BMVA).

Levin, A., Weiss, Y., Durand, F., and Freeman, W. T. (2009). Understanding and evaluating blind deconvolution algorithms. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1964–1971. IEEE.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International journal of computer vision*, 30(2):79–116.

Liu, R. and Jia, J. (2008). Reducing boundary artifacts in image deconvolution. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 505–508. IEEE.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee.

Lyons, A. P. and Abraham, D. A. (1999). Statistical characterization of high-frequency shallow-water seafloor backscatter. *The Journal of the Acoustical Society of America*, 106(3):1307–1315.

Madureira, L., Sousa, A., Sousa, J., and Gonçalves, G. (2009). Low cost autonomous underwater vehicles for new concepts of coastal field studies. *Journal of Coastal Research*, pages 238–242.

Mallios, A., Ridao, P., Ribas, D., Carreras, M., and Camilli, R. (2015). Toward autonomous exploration in confined underwater environments. *Journal of Field Robotics*.

Marston, T. M. and Kennedy, J. L. (2016). Volumetric acoustic imaging via circular multipass aperture synthesis. *IEEE Journal of Oceanic Engineering*, 41(4):852–867.

Martin, W. N. and Aggarwal, J. K. (1983). Volumetric descriptions of objects from multiple views. *IEEE transactions on pattern analysis and machine intelligence*, (2):150–158.

Masuda, T., Sakaue, K., and Yokoya, N. (1996). Registration and integration of multiple range images for 3-d model construction. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 1, pages 879–883. IEEE.

Matsebe, O., Kumile, C., and Tlale, N. (2008). A review of virtual simulators for autonomous underwater vehicles (auvs). *IFAC Proceedings Volumes*, 41(1):31–37.

McKinney, C. M. and Anderson, C. (1964). Measurements of backscattering of sound from the ocean bottom. *The Journal of the Acoustical Society of America*, 36(1):158–163.

Meagher, D. (1982). Geometric modeling using octree encoding. *Computer graphics and image processing*, 19(2):129–147.

Mizuno, K. and Asada, A. (2014). Three dimensional mapping of aquatic plants at shallow lakes using 1.8 mhz high-resolution acoustic imaging sonar and image processing technology. In *Ultrasonics Symposium (IUS), 2014 IEEE International*, pages 1384–1387. IEEE.

Murino, V., Trucco, A., and Regazzoni, C. S. (1998). A probabilistic approach to the coupled reconstruction and restoration of underwater acoustic images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):9–22.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234.

Newman, P. and Durrant-Whyte, H. (1998). Using sonar in terrain-aided underwater navigation. In *Robotics and automation, 1998. Proceedings. 1998 IEEE international conference on*, volume 1, pages 440–445. IEEE.

Nygren, I. and Jansson, M. (2004). Terrain navigation for underwater vehicles using the correlator method. *IEEE Journal of Oceanic Engineering*, 29(3):906–915.

Osfield, R., Burns, D., et al. (2004). Open scene graph. *Library–OSG. http://www. openscenegraph. org*.

Ozog, P., Troni, G., Kaess, M., Eustice, R. M., and Johnson-Roberson, M. (2015). Building 3d mosaics from an autonomous underwater vehicle, doppler velocity log, and 2d imaging sonar. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 1137–1143. IEEE.

Pailhas, Y. (2013). *Sonar systems for object recognition*. PhD thesis, Heriot-Watt University.

Palomer, A., Ridao, P., and Ribas, D. (2016). Multibeam 3d underwater slam with probabilistic registration. *Sensors*, 16(4):560.

Papadopoulos, G., Kurniawati, H., Shariff, A. S. B. M., Wong, L. J., and Patrikalakis, N. M. (2011). 3d-surface reconstruction for partially submerged marine structures using an autonomous surface vehicle. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3551–3557. IEEE.

Petillot, Y., Reed, S., and Bell, J. (2002). Real time auv pipeline detection and tracking using side scan sonar and multi-beam echo-sounder. In *OCEANS'02 MTS/IEEE*, volume 1, pages 217–222. IEEE.

Petillot, Y., Ruiz, I. T., and Lane, D. M. (2001). Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar. *IEEE Journal of Oceanic Engineering*, 26(2):240–251.

Pfingsthorn, M., Birk, A., and Vaskevicius, N. (2011). Semantic annotation of ground and vegetation types in 3d maps for autonomous underwater vehicle operation. In *OCEANS 2011*, pages 1–8. IEEE.

Prats, M., Pérez, J., Fernández, J. J., and Sanz, P. J. (2012). An open source tool for simulation and supervision of underwater intervention missions. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 2577–2582. IEEE.

Reed, S., Petillot, Y., and Bell, J. (2003). An automatic approach to the detection and extraction of mine features in sidescan sonar. *IEEE Journal of Oceanic Engineering*, 28(1):90–105.

Reed, S., Petillot, Y., and Bell, J. (2004). Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information. *IEE Proceedings-Radar, Sonar and Navigation*, 151(1):48–56.

Roman, C. and Singh, H. (2004). Micro-bathymetric mapping using acoustic range images. In *OCEANS'04. MTTS/IEEE TECHNO-OCEAN'04*, volume 3, pages 1574–1579. IEEE.

Roman, C. and Singh, H. (2005). Improved vehicle based multibeam bathymetry using sub-maps and slam. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3662–3669. IEEE.

Roman, C. N. (2005). *Self consistent bathymetric mapping from robotic vehicles in the deep ocean*. PhD thesis, Citeseer.

Rosenblum, L., Kamgar-Parsi, B., Belcher, E., and Engelsen, O. (1991). Acoustic imaging: The reconstruction of underwater objects. In *Proceedings of the 2nd conference on Visualization'91*, pages 94–101. IEEE Computer Society Press.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE international conference on*, pages 2564–2571. IEEE.

Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268.

Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the icp algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 145–152. IEEE.

Rusu, R. B., Blodow, N., and Beetz, M. (2009). Fast point feature histograms (fpfh) for 3d registration. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3212–3217. Citeseer.

Rusu, R. B. and Cousins, S. (2011). 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1–4. IEEE.

Sac, H., LEBLEBİCİOĞLU, M. K., and AKAR, G. (2015). 2d high-frequency forward-looking sonar simulator based on continuous surfaces approach. *Turkish Journal of Electrical Engineering & Computer Sciences*, 23(Sup. 1):2289–2303.

Sæbø, T. O., Synnes, S. A. V., and Hansen, R. E. (2013). Wideband interferometry in synthetic aperture sonar. *IEEE Transactions on Geoscience and Remote Sensing*, 51(8):4450–4459.

Saucan, A.-A., Sintes, C., Chonavel, T., and Le Caillec, J.-M. (2015). Model-based adaptive 3d sonar reconstruction in reverberating environments. *IEEE Transactions on Image Processing*, 24(10):2928–2940.

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280.

Schock, S. G. (2004). Remote estimates of physical and acoustic sediment properties in the south china sea using chirp sonar data and the biot model. *IEEE Journal of Oceanic Engineering*, 29(4):1218–1230.

Shan, Q., Jia, J., and Agarwala, A. (2008). High-quality motion deblurring from a single image. In *Acm transactions on graphics (tog)*, volume 27, page 73. ACM.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.

Shi, J. et al. (1994). Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE.

Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., and Webb, R. (2016). Learning from simulated and unsupervised images through adversarial training. *arXiv preprint arXiv:1612.07828*.

Slawski, M., Hein, M., et al. (2013). Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. *Electronic Journal of Statistics*, 7:3004–3056.

Snellen, M., Siemes, K., and Simons, D. G. (2011). Model-based sediment classification using single-beam echosounder signals. *The Journal of the Acoustical Society of America*, 129(5):2878–2888.

Soloviev, A., Maingot, C., Agor, M., Nash, L., and Dixon, K. (2012). 3d sonar measurements in wakes of ships of opportunity. *Journal of Atmospheric and Oceanic Technology*, 29(6):880–886.

Stanic, S., Briggs, K., Fleischer, P., Ray, R., and Sawyer, W. (1988). Shallow-water high-frequency bottom scattering off panama city, florida. *The Journal of the Acoustical Society of America*, 83(6):2134–2144.

Steder, B., Rusu, R. B., Konolige, K., and Burgard, W. (2011). Point feature extraction on 3d range scans taking into account object boundaries. In *Robotics and automation (icra), 2011 ieee international conference on*, pages 2601–2608. IEEE.

Stone, W. C., Ende, B. A. a., Wefer, F. L., and Jones, N. A. (2000). Automated 3d mapping of submarine tunnels. In *Robotics 2000*, pages 148–157.

Sun, N., Shim, T., and Cao, M. (2008). 3d reconstruction of seafloor from sonar images based on the multi-sensor method. In *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, pages 573–577. IEEE.

Thorner, J. E. (1990). Approaches to sonar beamforming. In *Southern Tier Technical Conference, 1990., Proceedings of the 1990 IEEE*, pages 69–78. IEEE.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Tikhonov, A. N., Goncharsky, A., Stepanov, V., and Yagola, A. G. (2013). *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media.

Trevorrow, M. V. (2004). Statistics of fluctuations in high-frequency low-grazing-angle backscatter from a rocky sea bed. *IEEE Journal of Oceanic Engineering*, 29(2):236–245.

Trucco, E., Fusiello, A., and Roberto, V. (1999). Robust motion and correspondence of noisy 3-d point sets with missing data. *Pattern recognition letters*, 20(9):889–898.

Urick, R. (1954). The backscattering of sound from a harbor bottom. *The Journal of the Acoustical Society of America*, 26(2):231–235.

Urick, R. J. (1967). *Principles of underwater sound for engineers*. Tata McGraw-Hill Education.

Vaneck, T. W., RODRIGUEZ-ORTIZ, C. D., Schmidt, M. C., and Manley, J. E. (1996). Automated bathymetry using an autonomous surface craft. *Navigation*, 43(4):407–419.

Wahl, E., Hillenbrand, U., and Hirzinger, G. (2003). Surflet-pair-relation histograms: a statistical 3d-shape representation for rapid classification. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 474–481. IEEE.

Wenz, G. M. (1962). Acoustic ambient noise in the ocean: Spectra and sources. *The Journal of the Acoustical Society of America*, 34(12):1936–1956.

Whitcomb, L. L. (2000). Underwater robotics: Out of the research laboratory and into the field. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 1, pages 709–716. IEEE.

Williams, S. and Mahon, I. (2003). A terrain-aided tracking algorithm for marine systems. In *Field and Service Robotics*, pages 93–102. Springer.

Woock, P. (2011). Deep-sea seafloor shape reconstruction from side-scan sonar data for auv navigation. In *OCEANS 2011 IEEE-Spain*, pages 1–7. IEEE.

Woodward, W. R., Webb, P. R., Hansen, R. K., Cunningham, B., and Markiel, J. (2010). Approaches for using three dimensional sonar as a supplemental sensor to gnss. In *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*, pages 1112–1117. IEEE.

Xi, F. and Shu, C. (1999). Cad-based path planning for 3-d line laser scanning. *Computer-Aided Design*, 31(7):473–479.

Yang, J., Li, H., and Jia, Y. (2013). Go-icp: Solving 3d registration efficiently and globally optimally. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1457–1464.

Yang, S. X. and Luo, C. (2004). A neural network approach to complete coverage path planning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):718–724.

Yufit, G. and Maillard, E. P. (2013). 3d forward looking sonar technology for surface ships and auv: Example of design and bathymetry application. In *Underwater Technology Symposium (UT), 2013 IEEE International*, pages 1–5. IEEE.

Zerr, B. and Stage, B. (1996). Three-dimensional reconstruction of underwater objects from a sequence of sonar images. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 927–930. IEEE.

Zimmerman, M. J. (2004). A 3d, forward looking, phased array, obstacle avoidance sonar for autonomous underwater vehicles.